The Dissertation Committee for Jennifer Chentzu Pai Certifies that this is the approved version of the following dissertation:

Engineering Peptide Specific Hyper-crystallizable Antibody Fragments (scFv) as Potential Chaperones for Co-Crystallization

Committee:

Jennifer A. Maynard, Supervisor

Hal Alper

George Georgiou

Thomas M. Truskett

Jon D. Robertus

Engineering Peptide Specific Hyper-crystallizable Antibody Fragments (scFv) as Potential Chaperones for Co-Crystallization

by

Jennifer Chentzu Pai, B.S., M.S.E.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2010

Dedication

I would like to dedicate this work

To my mother and father

To my loving fiancé, Jonathan

Acknowledgements

First, I would like to thank my professor, Dr. Jennifer A. Maynard, for allowing me to work on this project. I am grateful for her words of wisdom, support and guidance during my graduate studies at the University of Texas at Austin.

I would like to acknowledge Dr. Raquel Lieberman (School of Chemistry and Biochemistry and Institute for Bioscience and Bioengineering, Georgia Institute of Technology) for their collaboration with us on the hypercrystallizable chaperone crystallization studies and the privilege of working with their membrane protein. I would like to thank two members of her lab: Dr. Jason Drury and Jeffrey Culver for their assistance and willingness to answer my questions. I am grateful to Dr. Whitney Yin, Dr. Jon Robertus and Dr. Art Monzingo for use of their crystallization facilities at UT Austin.

Acknowledgement also goes to my current and former colleagues for their help and words of advice: Dr. Jamie Sutherland, Ryan Myhre, Benjamin Roy, Tarik Khan, Christine Chang, Kevin Entzminger, Zachary Frye. A special thanks goes to the graduate students and undergraduates who assisted me on different portions of the work in this dissertation: Jeong-min Hyun, Alan Goran, Kevin Lu, Carolyn Smith and Kim Yen.

Lastly but most importantly, much gratitude goes to my family: my parents for their love and support as I accomplished this goal, my sister and brother for their willingness to listen to my struggles. Much love goes to Jonathan for enduring with me through all the challenges and always believing in me.

Engineering Peptide Specific Hyper-crystallizable Antibody Fragments (scFv) as Potential Chaperones for Co-Crystallization

Publication No._____

Jennifer Chentzu Pai, PhD

The University of Texas at Austin, 2010

Supervisor: Jennifer A. Maynard

Hydrophobic membrane proteins perform a variety of important functions in the cell, but their structures are notoriously difficult to solve. Thus, new strategies to obtain crystals of membrane proteins for structure determination are critical. We aim to develop a toolbox of peptide specific single-chain antibody fragment chaperones engineered for hyper-crystallizability. These peptide sequences can be introduced into various regions of membrane proteins without interfering with protein function. The resulting protein-chaperone complex is expected to form a crystal lattice mediated by chaperone interactions.

We have developed candidate scFv chaperone proteins binding hexa-histidine (His_6) and EYMPME (EE) tags with improved biophysical features influencing crystallization propensity, including peptide affinity, stability and solubility. The scFv libraries were generated using a novel ligation-free technique, MegAnneal, allowing us to rapidly generate large libraries based on 3D5 scFv. We identified two candidate

chaperones, 3D5/His_683, specific for His_6 and 3D5/EE_48, specific for EE tags. Variants exhibit high solubility (up to 16.6 mg/ml) and nanomolar peptide affinities; complexes of 3D5/EE_48 with EE-tagged proteins were isolated by gel filtration. We have developed design rules for EE peptide placement at terminal, inter-domain or internal loop regions of the target protein to balance peptide accessibility for chaperone binding while retaining rigid protein-chaperone complexes suitable for crystallization.

The 3D5/ His_683 crystallized in four different conditions, utilizing multiple space groups. The 3D5/EE_48 scFv was crystallized (3.1 Å), revealing a ~52 Å channel in the crystal lattice, which may accommodate a small peptide-tagged target protein. Our evolution experiments altered scFv surface residues, resulting in use of different crystallization contacts. Analysis of these crystal contacts and those used by crystallized 14B7 scFv variants, led us to postulate that lattice formation is driven by strong crystal contacts. To test this hypothesis, we introduced amino acid changes expected to reduce the affinity of the 3D5/EE_48 energetically dominant crystal contacts. This approach to crystal contact engineering may allow semi-rational control over lattice networks preferred by scFv chaperones. Co-crystallization trials with model proteins are on-going. These engineered scFvs represent a new class of chaperones that may eliminate the need for de novo identification of candidate chaperones from large antibody libraries.

Table of Contents

List of Tables

List of Figures

CHAPTER ONE: Introduction & Background

1.1 PROTEIN CRYSTALLIZATION

1.1.1 The value & applications of crystallization

Crystallization of proteins has three major applications: i) bioseparation processes; ii) controlled drug delivery systems and most importantly; and iii) structural determination. The first two applications are driven primarily by the Pharmaceutics industry. In bioseparation processes, crystallization is commonly employed for purifying a product from the other biomass during downstream production (ex. Lysozyme, Insulin). This form of separation is advantageous because it does not cause unfolding or loss of activity [1]. Controlling the crystallization process and improving current crystallization unit designs are the major challenges involved. The potential use of crystals as a drug delivery system (such as crystalline lysozyme [2, 3]) focuses largely on producing crystals of uniform size to allow administering controlled doses over extended periods of time. While both of these first two applications are critical and have driven the research for improved protein crystallization, structural determination has played a major role in advancing our knowledge of protein functionality.

Structural determination of proteins at the atomic level is important for providing insight into the mechanisms of protein function, which guide the fields of protein engineering, vaccine development and drug design. Currently the optimal method for obtaining three-dimensional, atomic resolution structural models for large proteins is through crystallization with x-ray crystallography. Obtaining structural models of integral membrane proteins are of the highest priority with the NIH announcing a new structural genomics initiative in conjunction with Protein Structure Initiative Structural

Genomics Knowledgebase [4]. Accounting for approximately one third of the proteins in a cell, integral membrane proteins are vital to maintaining cell function with a myriad of responsibilities. These include, but are not limited to, i) transport molecules such as ions, water peptides and drugs across electrochemical gradients; ii) receive and relay signals from the extra- to the intra-cellular domains (or between cells); iii) provide openings (via channels) for molecules to move across membranes by means of diffusion. Their various roles make membrane proteins potential targets for drug delivery. Currently the largest category of drug targets (~50%) is directed at one particular class of membrane proteins: G-protein-coupled receptors (GPCRs) [5, 6]. The little structural information for these proteins has limited drug design to approaches guided primarily by protein mechanism. Incorporating a three-dimensional visualization of drug targets (including conformational intermediates) can clarify the role of protein structure in health and disease resulting in a organized and systematic approach to drug design. Establishing a general engineering approach to producing crystalline proteins [7, 8] can help to expand the limited structure we have. The ability to attain these visualizations has guided the motivation for improved crystallization processes. Understanding of the crystallization process will aid in the structural genomics project as well as expand its reach into bioseparation processes and controlled drug delivery.

1.1.2 The current challenges of protein crystallization

Although advances in NMR have achieved structural visualization of small protein molecules [9], in order to acquire atomic resolution images of large molecules (up to 150kDa) a crystal form is required. To obtain these crystal forms, several challenges need to be overcome, including i) production and purification of adequate amounts of

soluble homogenous proteins; ii) biophysical characteristics – soluble with limited hydrophilic or polar surfaces; and iii) determination and optimization of crystallization conditions [10-12]. Not only do crystals have to be produced, but they must be of high quality suitable for single crystal x-ray diffraction. Based on the number of targets deposited by worldwide contributing centers in TargetDB, of the 269,056 successfully cloned proteins, only 9.1% have currently been crystallized, 4.6% produced crystals large enough and suitable for x-ray of neutron diffraction studies with 3.6% producing three-dimensional crystal structures from the diffraction data (Table 1). The low yields are the result of difficulties that arise in part to the structural flexibility of proteins in solution and its ability to form many different intermolecular contacts based on slight changes in pH, ionic strength, temperature, protein concentration or specific additives in the crystallizing solution. While crystallization technique have successfully been used to determine ~60,000 protein structures, only ~100 of these are membrane proteins [Protein Data Bank]. Membrane proteins in particular are especially challenging due to their hydrophobic/amphiphilic characteristics, requiring additional detergents to prevent aggregation and aid in crystallization [13].

Table 1.1 Success rate of protein crystallization based on TargetDB status statistics for SG Centers submissions

Status	Total Number of Proteins	(%) Relative to "Cloned" Proteins
Cloned	128675	100.0
Expressed	86134	66.9
Soluble	33652	26.2
Purified	29954	23.3
Crystallized	10853	8.4
Diffraction-quality Crystals	5471	4.3
Diffraction	4932	3.8
Crystal Structure	4002	3.1
NMR Structure	1690	1.3
Work Stopped	33638	-

Reference: http://targetdb.pdb.org/statistics/TargetStatistics.html
(Last updated: September 24, 2008)

1.1.3 The mechanics of crystallization: the kinetics and thermodynamics

High diffraction quality crystals are formed when the proteins are immobilized within the lattice in a consistent formation (repeating units of identical orientation). Crystal growth rate is dependent on both kinetics and thermodynamics with three stages of growth: i) nucleation ii) post-nucleation and iii) cessation of growth [14] (Fig 1A). The growth of the crystal is dependent on the nucleation stage, because the quality of the critical nucleus typically drives the entire crystallization process. The initial driving force is thermodynamic in nature with an energy barrier and entropic in origin, which must be surpassed to form the initial nucleus. Crystal growth is driven by changes in Gibb's free energy, which is dependent on the entropies of the protein and the solution [15]. The equations below define the thermodynamic drive involved in protein crystallization, where $\Delta G^o_{cryst} < 0$ is a thermodynamically favored process.

$$\Delta G^o_{cryst} = \Delta H^o_{cryst} - T\Delta S^o_{cryst}, \quad \text{(Eq 1.1)}$$

$$\Delta G^o_{cryst} = \Delta H^o_{cryst} - T\left(\Delta S^o_{protein} + \Delta S^o_{solvent}\right)_{cryst} \quad \text{(Eq 1.2)}$$

$$K_{cryst} = \exp\left(-\frac{\Delta G^o_{cryst}}{RT}\right) \quad \text{(Eq 1.3)}$$

$\Delta S^o_{protein}$ is dependent on both the loss of the molecular degrees of freedom in addition to the loss of conformational freedom from the amino acids at the intermolecular contact interfaces. The extent of supersaturation (i.e. maintained thermodynamic equilibrium) is the driving force behind the phase transitions of a protein, which includes crystallization. During the post nucleation stage, crystal growth is driven largely by diffusion of molecules to the nucleus and surface kinetics at the edges of the crystal.

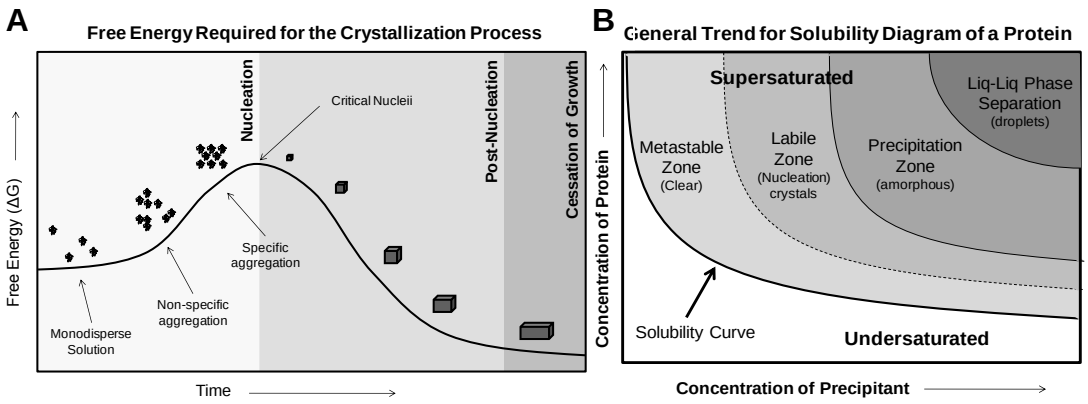


Figure 1.1 Protein Crystallization Diagrams

(a) Free energy diagram involved with the three stages of protein crystallization i) nucleation, ii) post-nucleation and iii) cessation of growth.
(b) An example of a solubility phase diagram for a protein.
Adapted from: http://www-structmed.cimr.cam.ac.uk/Course/Crystals/Theory/phases.html

The extent of supersaturation (i.e. maintained thermodynamic equilibrium) is the driving force behind the phase transitions of a protein which includes crystallization (Fig. 1.1B). During the post nucleation stage, crystal growth is driven largely by diffusion of molecules to the nucleus and surface kinetics at the edges of the crystal. While focus of protein crystallization has centered largely on the kinetics and thermodynamics of the first two stages, all three stages are necessary in producing diffraction quality crystals. As a result different crystal growth optimization approaches are currently in practice focus on each stage.

1.1.4 Roles of interactions and intermolecular contacts in protein crystallization

Although scientists are split on whether kinetics or thermodynamics intiates the crystallization process, it is clear that crystal contacts play a large role. As an attempt is

made to develop a solution to the crystallization dilemma, protein-protein interactions will need to be taken into account. It has become clear that geometrically precise intermolecular contacts formed with surface epitopes are required for lattice stability over amorphous aggregation [16, 17]. While there still remains no clear marker of crystalline ability that can be inferred from the primary sequence level, there have been numerous groups that have attempted to develop software to predict a protein's propensity to crystallize based on its primary sequence [18, 19]. The major disadvantages to this prediction software are that it relies on datasets of previously crystallized protein, and is limited to prediction of small/medium sized non-homologous proteins.

As a result, the nature of the intermolecular interaction of the proteins, as well as the mechanisms and assembly kinetics for the protein into an ordered crystal lattice is of specific interest. McPherson aptly puts the dilemma: "*The major challenges that remain in this area are to understand what kinds of mutation are most valuable in increasing solubility, enhancing lattice contacts, and discouraging unfavorable interaction, and then to employ this knowledge to predict useful modifications [20].*" Different interactions are strongly thought to correlate with protein-protein interaction including surface hydrophobicity [21], surface electrostatic potential of the protein [22] and a combination of surface electrostatic and hydrophobic interactions [23, 24]. We propose to find correlations between the biophysical characteristics of the protein and the resulting intermolecular contacts.

The Young Modulus for protein crystal displays the relative importance of intermolecular contacts and its impact on the crystal structure and role in defining the properties of the crystals [25, 26]. Here the small areas of weakly bound contact explain the lower Young moduli that are seen for protein crystals. Intermolecular contacts within the crystal lattice are shown to consist of patches that cover only a small percentage of

the surface of proteins. The patches combined cover approximately less than half of the total protein surface area. In many cases, the interacting patches consist of complimentary polar (H-bonds and salt brides) and charged side chains as well as van der waals attractive forces (partial electric charges on amino acids). While these lattice constants are biologically nonspecific, the forces guiding these interactions are similar to those for protein folding and protein-protein binding specificity. The variety of intermolecular contacts that encompass the entire surface of the lysozyme confirms that any molecular surface of the protein can potentially form lattice contacts depending on the conditions of crystal growth [26, 27]. At present, there is no rationale for determining conditions that can activate potential molecular patches for intermolecular interactions and crystal packing.

1.1.5 Characterization of protein interactions

Different interactions are strongly thought to correlate with protein-protein interaction including surface hydrophobicity [21], surface electrostatic potential [22] and a combination of surface electrostatic and hydrophobic interactions [23, 24]. Each of these interactions focuses on the protein and solution that it is immersed in. The majority of proteins have rather complicated non-spherical shapes and in many cases the proteins are located loosely in the crystal lattice in order to accommodate water and/or electrolytes in the cavity. Electrolytes surrounding the proteins in solution can induce crystallization and other solid-phase formations by controlling the like-charged protein molecules. A theory based of this concept known as the Derjaquin, Landua, Verwey & Overvel (DVLO) Theory of Colloid Stability focuses on electrolytes in the solution mediating van der waals attractions and electrostatic repulsions between the protein molecules [28,

29]. The hydrophobic attraction during crystallization is further enhanced also by the electrolytes while hydration repulsion occurs from a result of hydrate ions built up at the protein molecule surface [25, 30]. Analysis of lysozyme in varying concentrations and salt conditions, suggest that specific attractive interaction correlated with the crystallization/precipitation behavior, which was reflected in the retardation of the protein rotational diffusion [31].

Trying to predict and engineer intermolecular contacts has taken two different routes: (i) manipulating the protein and its solution or (ii) modifying the surface biophysical characteristics of the protein to improve the intermolecular contact. Molecular Interactions within protein solutions have been studied using static and dynamic light scattering (DLS) for analysis of the second virial coefficient. It has been found that the lower limit of B_2 corresponds to a low average protein-protein attraction in solution while a high limit results in aggregation [25]. The second virial coefficient is written as:

$$B_2(T) = 12 \int_0^{\infty} \left[1 - exp\left(\frac{U(\tilde{r},T)}{k_B T}\right)\right] \tilde{r}^2 d\tilde{r}, \qquad \tilde{r} = \frac{r}{2a} \qquad \text{(Eq 1.6)}$$

Where U(r) is the interaction potential between the molecules [32] and has been shown to have an interrelation with solubility via the interaction energy in the crystal. The highly negative values of B_2 will guide the selectivity of the crystallization process by allowing only specific molecular patches to collide and interact, thereby guiding the production of crystals that are oriented in the same direction. Aside from the osmotic virial coefficient mentioned above, other coefficients that have been used to study pairwise interactions and predict intermolecular contacts for crystallization include but are not limited to coefficients via sedimentation equilibrium [33], kinetic diffusion

coefficients [34] and optical scattering [35]. In addition to studying the biophysical characteristic of the protein surface epitopes, protein-protein interaction energies have been a point of interest. Studies have indicated that total interaction energy is fairly consistent between different proteins in their crystal lattice and this range needs to be retained for crystallization to continue [36].

Advances in recombinant techniques for crystallization and surveys of experimental results from the Northeast Structural Genomics Consortium, have led to the conclusion that surface epitopes drive the propensity for crystallization by providing intermolecular interactions [37]. Numerous physical properties were investigated to determine their control in protein crystallization including, but not limited to, thermodynamic stability and protein surface properties. The survey suggested that protein crystallization is dependent on the well-ordered intermolecular contacts rather than influence by its overall thermodynamic stability. In addition, it provides proof that low-entropy, well ordered surface features determine the behavior of the crystallization process. Work is still required to understand the nature of intermolecular interactions, the correlations of the biophysical characteristics, the mechanisms and assembly kinetics involved in a crystal lattice formation.

From the literature it appears that successful crystallization requires consistent intermolecular contacts, spatial orientation and organization. The challenge still remains of being able to predict a protein that has strong stereo-restricted interactions without promoting surface areas that can interact with multiple epitopes resulting in random aggregation. Current work in specific domain-domain interactions has allowed examination of interaction pairs and determination of interaction potentials [38]. While lattice contacts are less specific than these interprotein interactions, we hope to use our

observations to further our knowledge of crystallization initiation and apply it to engineering crystal lattice contacts.

The 14B7 will be utilized as a model protein for understanding the role intermolecular contacts play in crystallization. The antibody binds to a protective antigen (PA) of *B. Anthracis* and has been thoroughly studied for enhanced affinity [39]. The 14B7 antibody was selected due to its ability to produce large yields of monomers and its rapid crystallization (~24 hours). 14B7 is also of interest because of its propensity to change crystal space groups with 10 mutations, while retaining the largest crystal contact [40, 41]. Utilizing this antibody will allow us to study the contact energetics involved in driving lattice formation.

1.2 Different Approaches to Improving Crystallization

Numerous techniques have been used to generate protein crystals that include, but are not limited, to drop under oil, microbatch, dialysis, liquid-liquid free interface diffusion, capillaries, gels and microfluidics [11, 42]. Vapor diffusion utilizing sitting or hanging drops is one of the most popular crystallization techniques due to the small volume sample size, simple set-up and ability to vary physical parameters during crystallization. Many of these methods require unique protein crystallization conditions found through trial and error. Advancements in robotic technology have helped to facilitate methods like vapor diffusion and drop under oil, however, some finesse is still required. This shotgun approach for crystallization makes this process more of an art form. Scientists are struggling to develop a more systematic approach to this process by studying the mechanism that guides a protein on their path to crystallization, thereby understanding 'The Science of Crystallization.'

1.2.1 Art of crystallization

The advantages to an empirical method are its ability to determine the initial and unique conditions for crystallization. Historically, crystallography has used a shot-gun approach, testing thousands of conditions in parallel to provide a starting point for optimizing the production of diffraction-quality crystals. The first protein crystallized using this technique was a sperm whale homolog of the human myoglobin because crystallization of the human variant was not successful [43, 44]. Most of the initial protein structures were found using this empirical method, with numerous conditions and use of homologs (other species variants) with conserved function. Homologs have been used successfully to produce three dimensional structures of proteins that otherwise failed to produce diffracting crystals using the shotgun approach. The use of homologs assumes optimal success when functional regions are conserved (sequences main > 70% similarity) and diversity occursonly at solvent exposed residues. However, some homolog sequences have < 15% similarity and prediction of surface amino acids is difficult without structural information. In addition, the effort involved in subcloning each gene, determining protein production conditions and optimal crystallization conditions makes this method tedious and unfavorable.

The advancements in robotic technology and microfluidics, high throughput capable labs can perform approximately 1,000 screens with nanovolume samples per protein of interest [10, 20, 45, 46]. The use of screens has now expanded to include determining compounds, like ligands, that can help stabilize proteins, creating a bridge between the art and science [47]. While these advancements have increased the number of crystal trials, it still remains a black box with no guarantee of increased crystal

production. It is unclear whether successful crystal production is dependent on understanding the properties of both proteins and reagents to predict the protein's behavior in crystallization conditions. From this, two systematic approaches to crystal growth are currently being undertaken the technique-based strategy, which focus on the crystallization conditions and the protein-based strategy, where the focus is on the proteins biophysical characteristics.

1.2.2 Science of crystallization – optimization of crystallization conditions

As a result of empirical methods, engineers are attempting to understand the science involved with crystallization. Systematic methodologies have been developed to assist in more efficiently defining prospective crystallization conditions than the empirical shot-gun approach. These methodologies include: i) light scattering for determination of second viral coefficients [48, 49], ii) neutron, x-ray and light scattering for presence of pre-nuclear aggregates; ii) atomic force microscopy for kinetic studies of surface crystal growth and determination of kinetic based coefficient [50]; iv) phase diagram for optimization of crystallization conditions; and v) interferometry for delineation of crystallization mechanism [8, 20, 46, 51, 52].

One technique-based systematic approaches utilizes microfluidics to analyze a protein's phase diagram under several different conditions to determine an optimal crystallization condition (salt:protein concentration ratio, temperature, etc.). These devices use liq-liq free interface diffusion to provide kinetic optimization of crystallization reactions [45] or phase diagram visualizations [51, 53]. Protein nanocrystallography a new emerging field uses a combination of advances nanotechnologies to improve the crystal initiation conditions. Pechkova and Nicolini

have developed the Langmuir-Blodgett (LB) and Langmuir-Shafer (LS) technology, two innovative approaches to creating a protein thin-film nano-template as initial nucleation sites for further crystal growth [54, 55]. The motivation for optimizing crystal conditions is centered on determining the fundamentals of inducing and controlling the nucleation stage, ideally finding a 'universal' nucleant to drive the entire crystallization process [10, 56].

1.2.3 Science of crystallization – optimization of the proteins of interest

Recombinant techniques and the homology method have displayed there is a fine line between specific and nonspecific interactions that drive protein crystallization. Observations from these methods indicate a relationship between amino acid sequence and biophysical parameters, which suggests that proteins should be considered a variable in crystallography [57]. The successful use of homologs indicates the importance of surface morphology as a driving force in crystallization. Homology has paved the way for approaching the challenges of protein crystallization by addressing protein modification with recombinant techniques. The use of recombinant technique has helped address many of the bottlenecks at each level of the crystallization process (Table 1.1). The use of recombinant proteins has assisted in advancing crystallization with dramatic increases in protein yield (via various vectors); solubility and stability (via critical mutations) and facilitation of purification (via peptide tags) [58-60]. Additionally, the use of recombinant techniques can facilitate overcoming the final hurtle of solving the diffraction pattern with techniques such as protein labeling with SeMet. Recombinant proteins are now engineered with the intention that their improved biophysical characteristics play a critical role during the crystallization process. Numerous

recombinant techniques have aimed at reducing flexible regions [61, 62], truncating the protein into sections that are stable [58] and even removing glycosylation sites or unpaired cysteines, which have tendency to promote aggregation [60]. Modifications at intermolecular interactions of water soluble and membrane proteins indicate the sensitivity of surface residues substitutions to the integrity of the crystal contact [62, 63]. Use of molecular engineering, using both directed and random techniques, has provided a new direction for protein crystallization and has established that modifying the protein is a viable path.

1.3 MOLECULAR ENGINEERING VIA DIRECTED RECOMBINANT TECHNIQUES FOR IMPROVED CRYSTALLIZATION

1.3.1 Recombinant work with intermolecular contacts at the amino acid level (mutagenesis at known/suspected intermolecular contacts)

Experimental work with hypercrystallizable lysozyme has supported the importance of intermolecular contacts as the driving force in crystallization. Lysozyme has been crystallized under numerous conditions resulting in varying lattice contacts that span the entire surface area of the protein. Different molecular surface areas are activated as lattice contacts by varying pH, temperature and crystallization conditions. The activation of these distinct interaction areas result in different polymorphs of lysozyme. Recombinant work with T4 lysozyme has yielded various mutants crystallized in 25 different crystal lattice forms and growth kinetics, where even a single surface residue mutation can yield different space group [64]. Recent developments in recombinant work have now focused on analyzing specific intermolecular contacts, where the contacts are manipulated to explore its impact on protein crystallization. A modified 14B7 antibody, M18, with 10 mutations resulted in a drastic change in lattice formation and

growth kinetics, requiring a longer time for crystal formation than the wildtype [40, 41, 64]. Site-directed mutagenesis techniques have shown impact on protein solubility by removing large hydrophobic patches with neutral electrostatic potential, reducing the random aggregation and creating more ordered protein-protein interaction [58, 59, 65].

Through these studies the effects of surface residues substitutions have been shown to impact the integrity of the crystal contact and crystal packing. Mutations of surface residues on the *Thermus thermophilus* aspartyl-tRNA synthetase-1 were shown to influence crystal growth, packing arrangement and crystal quality [63]. The amino acids selected for modification were involved in lattice packing contacts and resulted in a correlation between protein crystallizability and manipulated packing interactions. From this study, it was observed that disruption of lattice contacts leads to poor or no crystal production while addition of potential contacts produced better crystals. This was further supported by a study with five single site mutants of *Rhodobacter sphaeroides*, where modified residues resided in contact regions that formed the tetragonal lattice. Results from the crystallization studies indicated that the mutations were significant for maintaining the stability of the tetragonal form. In addition, the loss of diffraction quality crystals indicated that the selected mutants carried specific interactions with neighboring residues within the contact [62].

1.3.2 Protein modification to reduce surface entropy

Site-directed mutagenesis on surface residues for rational protein crystallization has become well established in protein crystallography [66, 67]. Current work using site-directed mutagenesis has effectively improved protein crystallization by removing large hydrophobic patches with neutral electrostatic potential, reducing the random aggregation

and creating more ordered protein-protein interaction [58, 59, 65]. One site directed mutagenesis approach, surface entropy reduction (SER), focuses on utilizing mutations to reduce the protein's conformational surface entropy, allowing for thermodynamically favorable crystal contacts. Removal of amino acids with high conformational entropies produces the *homogeneous, contact forming surface patches* [15, 68] that are beneficial for crystallization. The thermodynamic cost of immobilizing high-entropy side chains tends to inhibit their participation in crystal packing contacts, by removing these side chains we reduce the cost involved [60]. The generation of these 'low-entropy' patches will reduce the cost requirement for burying surface regions into the crystal contact. This can be accomplished through removal of residues with high conformational energy and substitution with smaller amino acids. It has been shown that amino acids lysine, arginine and glutamate do not frequently appear in lattice contacts for this reason [69]. Further analysis supports the idea that successful utilization of this technique requires alanines substitutions in pairs or triplets to improve intermolecular contacts for crystallization. Variants of RhoGDI using these clustered mutations, successfully yielded higher order and diffraction quality crystals with this method [60], leading to the release of a SER web server for the design of crystallizable variants with clustered mutations to reduce amino acids with high conformational entropy [70].

While manipulating select residues is one of the more popular ways of reducing surface entropy, this concept has been extended to include strings of amino acids. Flexible termini of proteins as well as large flexible loops have also been targeted to reduce entropic impediment to crystallization [71]. Removal of these regions will minimize the heterogeneity of the protein as well as reduce the multiple varying interactions that can occur. Successful crystallization of the HIV gp120 envelope glycoprotein, shows the impact the replacement of two flexible loops with Gly-Ala-Gly

linkages can contribute [72, 73]. This strategy is especially beneficial for membrane proteins that characteristically contain large flexible intracellular loops. The replacement of these large loops with smaller linkers can produce crystals with higher resolution [74]. Considering that many membrane proteins contain at least one region of amino acids with no rigid secondary structure, removal of these regions will be the first step towards achieving improved membrane protein crystallization.

1.3.3 Introduction of contact forming regions

SER improves the thermodynamic state of the crystallization process, while other site directed mutagenesis approaches are even more systematic for generating crystal contacts. Modulations of intermolecular interactions indicate that crystallization is sensitive to surface residues substitutions [62, 63]. Novel methods have proposed to facilitate particular lattice formation with the creation of intentional interprotein contacts (Fig. 1.2). This is accomplished by introducing a small modified region on the target protein that will guide and possibly control the extent of crystallization. Banatao proposed the implementation of single unpaired cysteines into the protein so that intentional dimmers can be generated for more uniform crystallization units [75]. The cysteines are strategically placed on the surface of the protein to create symmetric proteins upon dimeration with a second protein (Fig. 1.2a). This creates a more uniform molecule, which will facilitate the ease of ordered crystal packing. The introduction of a leucine zipper-like hydrophobic interface (4 Leucines) into human RNase1 resulted in a designed hydrophobic packing crystal contact (Fig. 1.2b) [76]. The leucine incorporation has shown potential use for promoting intermolecular symmetry for crystallization and controlled crystal packing.

While small regions of manipulations are the most common method for utilizing this approach, a novel approach has been to fuse an additional whole protein to the protein of interest to assist in creating contact forming regions. Fused normally at the n- or c-terminal region, the ability to create these fused proteins is fairly straight forward. Fusion proteins have successfully been used for protein purification purposes with many well established vectors (pMal, NEB). The successful crystallization of the fused maltose binding protein (MBP)-target protein shows the potential of this strategy as a general approach to crystallization [77]. The additional MBP provides the lattice contacts necessary for crystal growth and serves as a chaperone, assisting in the crystal scaffold development (Fig. 1.2c). Recent work has advanced the use of fused proteins from insertion at the termini to engineering the "hyper-crystallizable" proteins in place of a flexible region on the membrane protein [61]. The ability of these proteins to remove the entropic cost contributed by the flexible loops as well as generate additional polar regions for lattice contacts is of particular interest in membrane protein crystallization. The GPCR membrane protein was successfully crystallized using this approach by inserting the hypercrystallizable T4 Lysozyme in place of one of the GPCR's flexible loops (Fig. 1.2d). The additional surface topography contributed by the T4 lysozyme provided the intermolecular contacts that drove its crystallization. While this method was successful, insertion of a second protein into an internal region of an existing protein without disturbing its structure or function can be especially challenging and time-consuming. Appropriate regions for insertion need to be selected to maintain proper folding of both the inserted chaperone protein as well as the protein of interest. For these reasons, this method's utility as a general approach to protein crystallization is questionable and may be as challenging as other crystallization techniques currently in place.

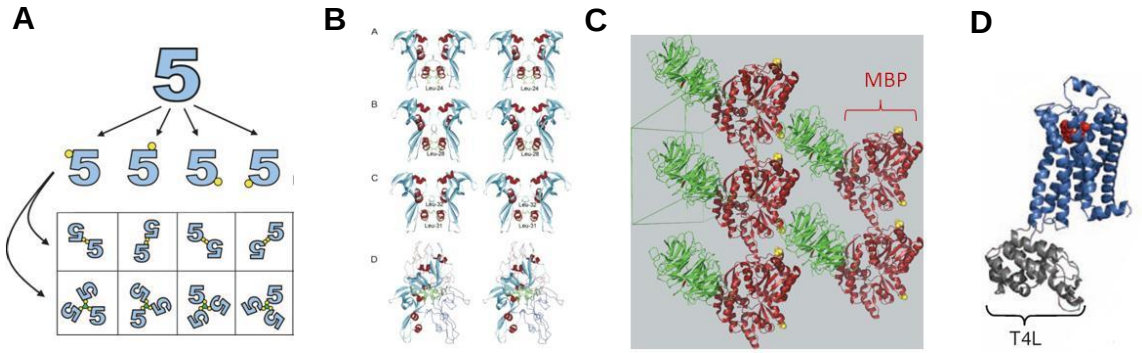


Figure 1.2 Different strategies to introducing intermolecular contacts for crystal lattice formation

(a) Directed cysteine insertions to create different symmetric units for crystallization [75]
(b) Insertion of Leucine zipper to create specific interactsion [76]
(c) Fusion of chaperone protein at terminal end of target protein (Chaperone: MBP in red, Target protein: RACK1 in green) [77]
(d) Fusion chaperone protein at internal loop of target protein (Chaperone: T4 Lysozyme in grey, Target Protein: β2-adrenergic receptor in blue) [61]

1.4 Molecular Engineering via Randomized Evolution - Uses in Crystallization

1.4.1 Gene shuffling between various homologs to improve

Besides site-directed mutagenesis, studies have indicated the potential of DNA shuffling [78]. Little rationale is involved and the diversity produced from shuffling homologs, modifies the surface properties enough to produce crystals. Another technique known as co-crystallization has been especially useful for crystallizing membranes proteins with large hydrophobic characteristics [79].

1.4.2 Co-crystallization: chaperone assisted scaffold

As we have mentioned above, there are numerous strategies to generate crystallizable proteins however, there still remains no generalizable approach. The use of

fused proteins (like MBP and T4 Lysozyme) as co-crystallizable chaperone serves primarily to introduce additional lattice contact regions. This technique has expanded to include protein specific recombinant proteins as chaperones that create stable complexes with improved crystallization properties [80]. The first case of co-crystallization was the successful use of an antibody chaperone in assisting crystallization [81]. Since then, this technique has assisted the crystallization of difficult proteins, in particular the membrane protein with their large hydrophobic regions. A significant portion of a membrane protein's surface is hydrophobic due to their integration within lipid bilayers. This characteristic poses a major dilemma as, once these proteins are removed from the bilayer, their tendency is toward nonspecific aggregation making them not readily soluble in aqueous solutions. The removal of these proteins can also compromise its physical stability leading to reduced or lost activity.

A number of approaches have previously been taken to overcome the hydrophobic characteristics dilemma and create the isotropic, monodisperse solutions necessary for protein crystallization. The most common method for improved crystallization involves tailoring the surfactant/detergent environment [79, 82]. Detergent monomers, serve as chaperone-like bodies and associate with the hydrophobic surfaces creating a "micelle-like" protein detergent aggregate. The resulting micelle-micelle interactions create additional phase transitions that can impact crystal nucleation and growth, which enhance nucleation and assist protein crystallization [82]. While more traditional methods of solubilized protein-surfactant complexes are more popular in membrane protein crystallization, a new and emerging areas involves the *in meso* approach. This technique employs a bicontinuous lipidic mesophase which attempts to build a lipid bilayer reservoir to assist in former crystal packing [13] and includes the *in cubo* approach, that employs lipidic cubic phases [83]. bicelle methods [84].

The use of recombinant proteins as an additional aid or chaperone for membrane protein crystallization works similarly to the prior detergent based methods. Both methods focus on countering the hydrophobic surfaces, however the protein specificity of the chaperone during complex formation allows for additional control. The chaperone protein recognizes and binds a membrane protein to form a stable complex. The resulting complex provides improved crystallization properties (i.e. more homogenous monomeric solution) by protecting the hydrophobic regions of the membrane protein and preventing numerous protein-protein interactions. These chaperones or co-crystallization proteins (CCP) can provide better probability for ordered crystals in two other mechanisms: (i) "lock downing" flexible regions to reduce heterogeneity and (ii) presenting additional polar surfaces available for crystal contact formation [7, 80, 85]. The polar regions provided by the chaperones produce lattice-forming crystal contacts that allow the POI to be suspended within the lattice, particularly beneficial for use with detergent solubilized proteins [79, 86]. Because the three-dimensional molecular structure of the chaperone is usually predetermined, this provides an added advantage as a molecular replacement model for solving the crystallographic phase and diffraction data of the complex.

Antibody fragments, Fab and Fv antibodies in particular, are currently the dominant chaperone in use for successful co-crystallization of membrane proteins, primarily due to their well characterized binding interactions [87] [86, 88, 89]. With the advancements of recombinant techniques however, efforts to generate alternative protein-binding chaperones are also in progress. Developments in this field have led to CCP protein options for co-crystallization which include but are not limited to affibodies [90], VHH Camelid Domains [85], and designed ankyrin repeat proteins (DARPins) [91, 92]. Below, we have shown the proteins that have been used successfully to co-crystallize target proteins (Fig. 1.3).

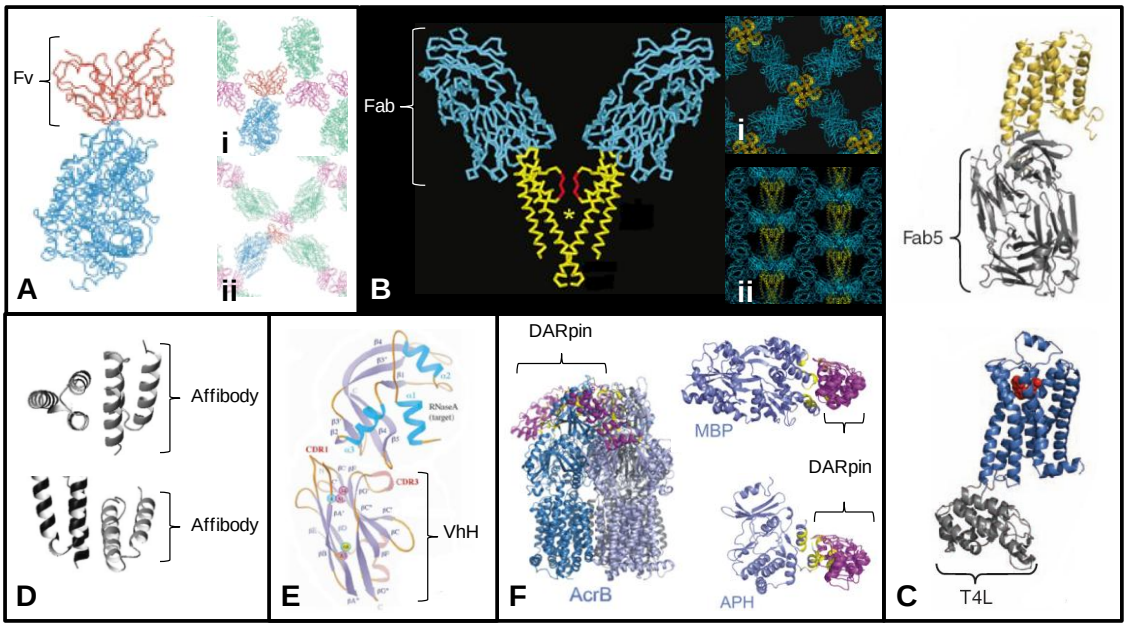


Figure 1.3 Structural images of successful co-crystallization proteins in complex with the target protein they helped crystallize

(a) Antibody Fv fragment – COX from P. denitrificans in complex [89]. i) crystal lattice of two-subunit, ii) crystal lattice of four-subunit;
(b) antibody Fab fragment - KcsA K+ channel [86]. i) view down four-fold axis of the I4 cell, ii) view perpendicular to the 4-fold axis;
(c) crystallization of gPCR using two different chaperones [61]. top: Fab5 – gPCR, bottom: T4L – gPCR;
(d) affibody – protein Z [92];
(e) VhH – RnaseA [85];
(f) co-crystallization of DARpins with AcrB, MBP and APH [92].

The recent chaperone-assisted breakthrough utilizing both Fab and Lysozyme chaperones independently to solve only the second known crystal structure of a GPCR (Fig. 1.3c) shows the effectiveness of co-crystallization [61, 93, 94]. The hydrophobic characteristics and the structural flexibility in one of its intracellular loops makes GPCRs unfavorable for crystallization and particularly challenging. The additional surface topography provided by both the T4 Lysozyme and the Fab assisted in mediating crystal contacts for crystallization. The successful crystallization with a high affinity Fab

highlights its negligible impact on the target proteins structure and the systematic approach for the protein-specific chaperone strategy for assisting crystallization. CCP antibodies have potential as a general application for protein crystallization if they can bind strongly and forming a rigid complex without affecting the native structure of the target tproteins. The in vitro selection of the Fab from a large library, shows a more direct approach to screening chaperones for assisted crystallization. However the search of a highly specific chaperone remains time-consuming. Not only does the chaperone need to be specific to the target protein it also must be characterized for crystallization propensity before it can become a successful chaperone. We propose the concept of a chaperone highly specific for a peptide tag, which can be added to any protein of interest. This peptide specific chaperone can provide a platform system that can reduce the time and effort that is currently invested in the present co-crystallization system.

1.5. ANTIBODIES FOR CHAPERONE ASSISTED CO-CRYSTALLIZATION

1.5.1 Structures and recombinant versions

To date, the leading CCPs in the field are antigen binding fragments (primarily Fabs-50kDa and Fv-25kDa). When complexed to proteins with aggregation properties, the biophysically favorable characteristics of the CCPs allow the resulting complex to become more soluble and monodisperse in solution [88]. One such antibody Fv fragment complexed with cytochrome c oxidase (COX), showed the antibody's ability to crystallize the complex into two different lattice conformations [89] (Fig. 1.3a). Examples of some natural and recombinant antibodies used in co-crystallization are shown below (Fig.1.4).

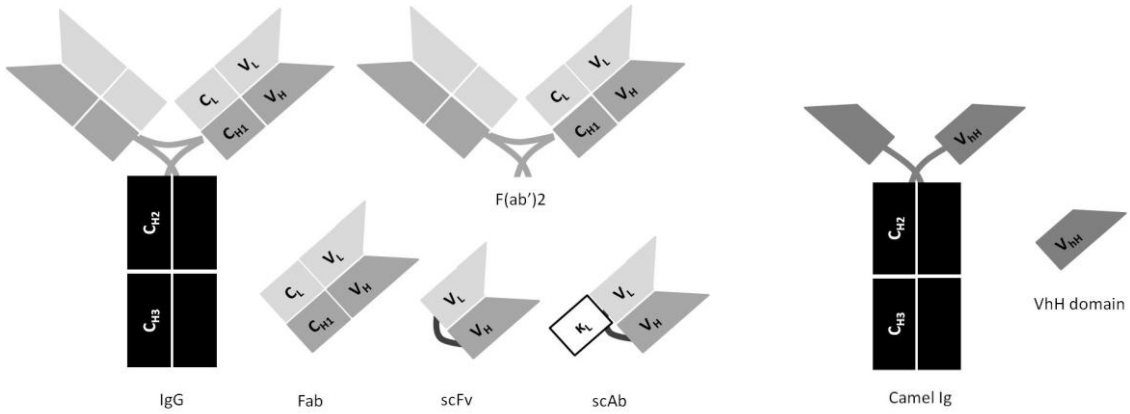


Figure 1.4 Natural and recombinant antibodies used in co-crystallization

Schematic drawing of different classes of antibodies and certain antigen binding fragments have been proteolytically cleaved and recombinantly developed from these classes.

Fabs contain a single constant chain (both light and heavy chain, C_L and C_H respectively) and a variable chain (V_L and V_H). The variable domain (V_L+V_H) consist of six hypervariable loops known as complementary determining regions (CDR) that bind the antigen with a specifity. Fab and Fv CCPs currently in use are produced via proteolytic cleavage from monoclonal antibodies [79], however advances in recombinant techniques have allowed for efficient production of these and even smaller antibodies. One particular small recombinant antibody commonly produced is the single chain variable fragment (scFv-27kDa) and is created by connecting the V_L and V_H domains with a flexible polypeptide linker [95]. Another single chain antibody variation involves attaching the light chain kappa constant domain to the scFv (scAb-43kDa) [96]. While some work has indicated the potential of scFv and scAb as co-crystal chaperone, more work is needed to determine the full extent of its capabilities [97]. The ability of these proteins to generate additional lattice contacts and provide a scaffold system is of particular interest and will guide the development of a crystallizable chaperone antibody.

1.5.2 Single chain antibody fragments for co-crystallization scaffold

The major limitation in current co-crystallization methods is the identification of a new chaperone with targeted affinity for each unique protein intended for crystallization. This challenge extends to selecting a chaperone that binds the "ideal" location on the protein without disturbing the native conformation [98]. Additionally, each new chaperone selected will have completely different biophysical characteristics. This results in screening of crystallization conditions for the new chaperone complexed with the target protein. To overcome the limitations of the current methods, a single chain fragment (scFv) antibody was engineered using established phage display technology for high affinity and specificity to a tag that could readily be attached to any protein of interest (via internal loop, or at either the c- or n-terminal ends). The scFv recombinant antibody is preferred for our objectives due to its compact nature, comprising a minimal binding unit: the variable light (V_L) chain domain and the variable heavy chain (V_H) linked together by a flexible $(Gly_4Ser)_4$ linker [95]. Hexahistidine-specific 3D5 scFv antibody [99-101] was selected as a candidate for use as the framework of the engineered antibody. Its selection was driven by its similar V_L homology (>85%) to a family of well-characterized antibodies that crystallize under various crystallization conditions (Fig. 1.5). In the 3D5 $P3_221$crystal lattice, the location of the crystal contacts on the antibody's framework makes this a potential and favored candidate for chaperone design.

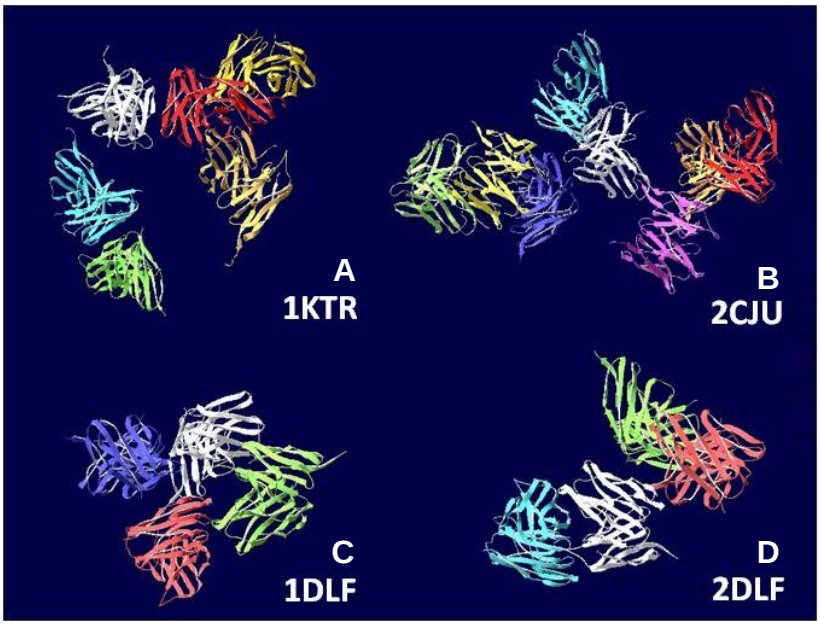


Figure 1.5. Crystalline Lattice Structure of Select Antibodies with similar V_L Homology to 3D5 scFv

(a) 3D5 scFv (Space Group: $P42_12_1$) [99]
(b) Anti-2-Phenyl-5-oxazolone NQ16-113.8 scFv (Space Group: $I2_12_12_1$) [102]
(c) Anti-dansyl IgG Fv fragment crystallized at pH 5.25 (Space Group: $P2_12_12_1$) [103]
(d) Anti-dansyl IgG Fv fragment crystallized at pH 6.75 (Space Group: $P2_12_12_1$) [103]

3D5 scFv is an anti-His antibody and currently the only one available in scFv form. A construct created by Kaufman consisting of nine mutations from the wildtype expressed extremely well and was capable of crystallizing with contacts along the framework as mentioned previously [99]. This allows for the manipulation or replacement of the CDRs for switched affinity for peptide tag while retaining the crystal contact. In addition to its other favored characteristics, 3D5 scFv was also selected for its large protein yield, crystallization at low protein concentrations and its crystal contacts are located on the frameworks. The 3D5 antibody interacts with only three c-terminal located histidines with a relatively weak binding affinity (K_D ~1uM). Its sensitivity to

pH, binding only between the range of 6 - 7.4, and poor solubility (<2.2mg/mL) requires that improvements need to be made to the antibody before it can be considered as a crystallization chaperone.

We propose to switch the affinity of the His-tag to an EYPME (EE)-tag. This tag was favored because of its improved chemical diversity in the peptide sequence, with a number of attractive residues i) a tyrosine, which frequently occur in protein-protein interactions and known for its hydrophobic interactions and hydrogen bond formation [104, 105] ii) a glutamate, negatively charged to form electrostatic interactions and iii) a proline that reduces the conformational entropy. A major proponent for selecting this peptide tag was that antibodies with specificity for the EE tag instead of the His tag have previously been established [106]. In addition, there are well behaved commercially available mAbs that have higher affinity and specificity for the EE peptide tag [Covance]. Further background and details on all of these antibodies and tags will be discussed in their respective chapters.

1.6 ANTIBODY ENGINEERING

1.6.1 Rational design

Directed evolution is a robust tool for understanding the relationship between protein structure and function. By introducing mutations into the gene that encodes the target protein, we can expand our limited understanding of protein mechanism and functionality. When structural information of the protein is present, Site directed mutagenesis (SDM) can be utilized to understand and improve protein activity. Use of targeted mutations at critical interaction, binding or active sites can increase the chances of obtaining an improved protein much more rapidly than with random mutagenesis (Fig.

1.6b). Although site directed mutagenesis (SDM) has gained in popularity, random and extensive mutagenesis (REM) is an effective tool for altering proteins for enhanced stability and improved activity [107, 108]. Two of the more commonly used REM techniques are gene shuffling and random insertion (Fig 1.6a). Some of the most popular applications for random mutagenesis involve improving antibody affinity to ligands [109-111], epitope mapping [112], function of proteins [113, 114], metabolic pathways and engineering promoters.

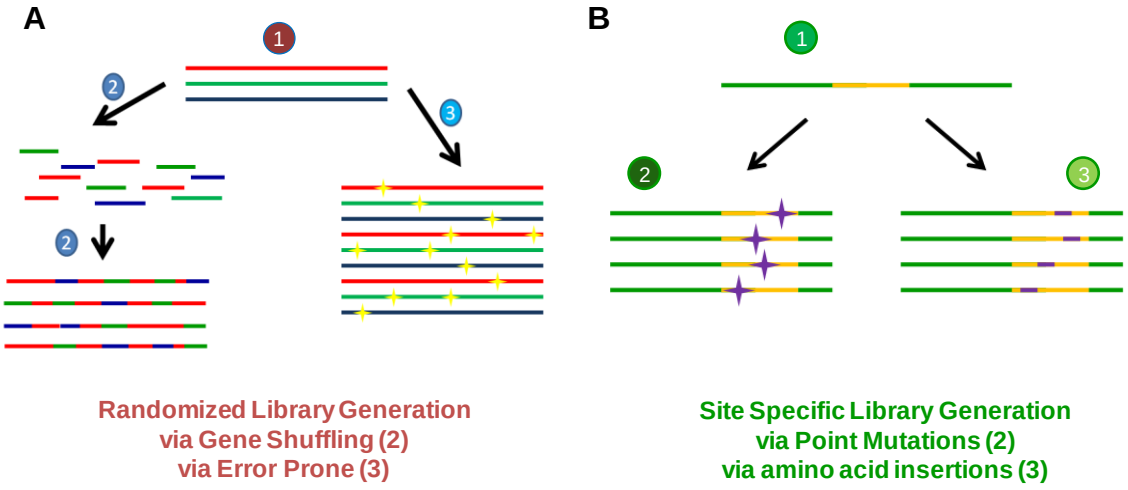


Figure 1.6 Two most frequently used methods for generation of protein libraries.

(a) Multiple genes used for gene shuffling (path 2) and random mutagenesis (path 3).
(b) Site directed mutagenesis utilizing switched amino acids (path 2) and inserted amino acids (path 3)

1.6.2 Methods for generating libraries to obtain appropriate chaperones

Recently, mutagenesis is more frequently accomplished with polymerase chain reaction (PCR)-based methods due primarily to the simplicity of these protocols. The

Stratagene QuikChange SDM Protocol (QCM) [115], one of the most widely used PCR-based methods for site directed mutagenesis, uses plasmid template and complementary synthetic oligonucleotides. The synthetic oligonucleotides contain the desired mutations and serve as primers to synthesize double stranded PCR product. Originally used for single mutations, recent variations of PCR methods sought to address and improve problems related to primer design and length. Some of these improved protocols include servers predicting successful amplifying oligonucleotides [116], partially overlapped oligonucleotides [117], overlap extensions with random mutations [118] or the production of megaprimers using non overlapped primers with an additional PCR cycle to complete whole plasmid double stranded synthesis [119-121]. The QuikChange method has also been expanded for application in REM by inducing random mutations during the around the plasmid PCR amplification step [122]. Stratagene's own variation: EZClone Domain Mutagenesis kit produces megaprimers using low fidelity polymerase (mutazyme) to initiate the around the plasmid PCR amplification. While the simplicity of the PCR methods is favorable, the exponential amplification of whole plasmid DNA allows for increased chances in errors at other locations on the plasmid (i.e. ori, promoter and antibiotic resistance) even with high fidelity polymerases such as pfu.

To avoid the potential errors introduced outside of the target area due to whole plasmid PCR, other protocols use error prone polymerase to amplify the regions for randomized mutations [123]. The randomized regions are then inserted into the vector of choice via ligation. With the ligation method, major challenges include i) efficient digestion of both the vector and insert ii) optimization of the ratio of insert to vector for successful ligation and iii) effective transformation into cells. Kunkel Mutagenesis with improvements from Weiss is another SDM technology that avoids PCR amplification and ligation limitations using template ss-DNA containing Uracils (U-DNA) [124]. Location

specific mutagenic primers anneal to the ss-DNA to covalently closed circular dna (heteroduplex). The U-DNA wildtype template becomes inactivated once transferred into ung^+ E. Coli, which allows for the strand containing the mutagenic megaprimer to be replicated. An inexpensive technique, the major advantage of this protocol is its ability to simultaneously anneal multiple independent primers at various regions for rapid library generation.

Other techniques avoid recombinant DNA work completely, using E. Coli mutator strains like DM2516 and XL1-Red (Stratagene). This method is extremely limited as the rate of mutations is extremely low and the mutations are uniformly inserted throughout the plasmid, impacting areas such as the promoter and antibiotic resistance [125]. Previous work in our lab with the XL1-red (deficient in mutS, mutD and mutT), has indicated the slow rate of mutation insertions into the plasmid (~10 – 6 mutations/kb/generation). After 15 rounds of cell growth, still no mutations were seen in our region of interest and confirmed previous work where few mutations were seen after 100 generations [125].

It is obvious from the current technologies in place that an improved protocol is needed for randomization. Pressure for better technologies involves overcoming the major bottlenecks from current SDM and REM protocols which include no longer relying on restrictions sites or ligase effectiveness, as well as reducing the amount of errors introduced outside the region of interest. Utilizing a combination of megaprimers and Weiss' improved Kunkel mutagenesis, we will present a novel method for random mutagenesis that can overcome the limitations of the current protocols currently in place for directed evolution.

1.6.3 In vitro selecion using phage display

In order to rapidly screen a library, an effective selection platform must be used. All library work in this thesis was performed using phage display selection techniques, a powerful and highly efficient technology for selecting proteins with desirable biological characteristics [126]. In the last 20 years, phage display systems have successfully been used to screen antibodies, ligands and other peptide-based molecules [127]. Known primarily for high affinity selection, phage display applications are wide-ranging and include selection for stabilized proteins [128], as well as improved structural folding and enzymatic activity [129]. The phage system in this thesis utilizes phagemid vectors with the gene inserted into the vector. Cells carrying this vector can be super-infected with helper phage to produce protein-presenting phage. These protein-presenting phage are incubated over bound target protein to find high binders (Fig. 1.7) eluted and the entire process repeated to improve specificity.

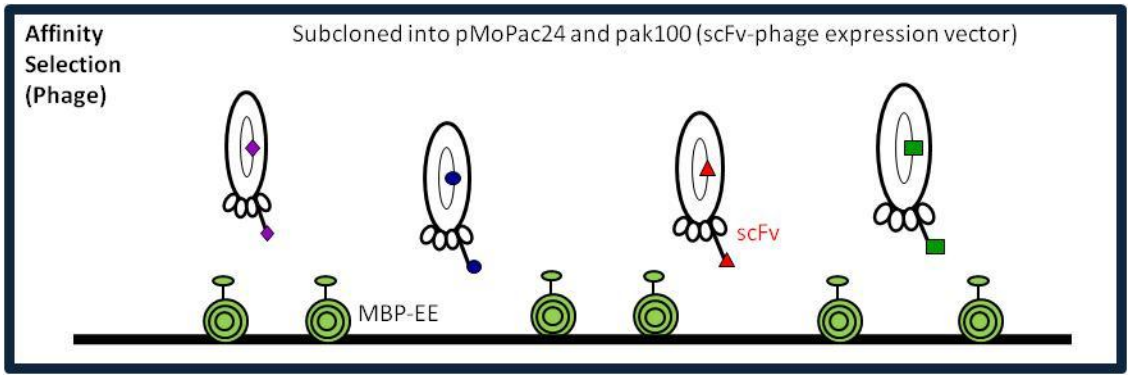


Figure 1.7 Affinity Selection for Phage Display

1.7 OBJECTIVES AND WORK

The thesis focuses on whether a hyper-crystallizable antibody could be engineered for potential use as a peptide specific chaperone. The objective is to create a hyper-crystallizable protein able to crystallize in multiple space groups and/or conditions with the intention of using it to nucleate crystal growth of more difficult protein. The development of a chaperone specific for a peptide tag would facilitate the path for a general approach by adding the cognate peptide to any membrane protein of interest at either terminal end or in place of a flexible loop. Use of this chaperone will simplify molecular replacement and increase the hydrophilic surface area available for forming lattice constants. By expanding this idea and applying it to other peptide tags, a more organized and systematic approach to protein crystallization is potentially viable (Fig. 1.8).

Here we report engineered 3D5 variants with switched peptide binding specificity and high affinity to the EE tag while retaining all the critical residues that driving crystallization. Variants were developed based on libraries with both directed and randomized evolution and resulted in a potential candidate for use as a crystal chaperone. While engineering this antibody, it was important to understand the correlation between its intermolecular contacts and crystallization propensity. This involves understanding the driving force behind lattice formation and if it can be controlled as we design our chaperone antibody. The impact of surface mutagenesis at each intermolecular contact is explored at an energetic level as well as its involvement within established crystal contacts.

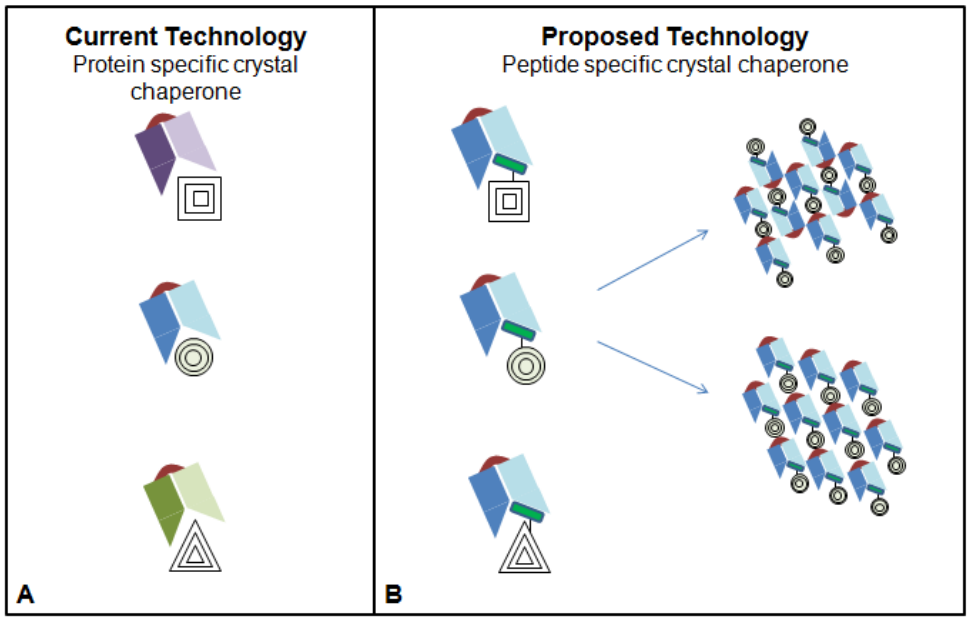


Figure 1.8 Chaperone-Assisted Protein Crystallization Systems

A general platform for a more organized and systematic approach to protein crystallization using peptide specific crystal chaperones in place of the currently existing protein specific crystal chaperone.

CHAPTER TWO: Restriction enzyme-free construction of random gene mutagenesis libraries in *E. coli*

2.1 CHAPTER SUMMARY

Directed evolution relies on both random and site-directed mutagenesis of individual genes and regulatory elements to probe the biochemical basis of protein activity and modify activity to meet engineering specifications. Central to these experiments is construction of large, unbiased libraries of related molecules; however, a number of technical hurdles continue to limit routine library construction. While technologies have been developed to avoid some or all of these challenges in yeast and for site-directed mutagenesis libraries in *E. coli*, no such methods exist for development of random mutagenesis libraries in *E. coli*. Here, we report a novel ligation-free library approach, based on error-prone PCR amplification of the DNA region of interest and use of this product to generate a library of randomly mutated 3' megaprimers, followed by megaprimer annealing to dUTP-containing, single-stranded template plasmid DNA. The T7 DNA polymerase then seamlessly integrates the modified DNA into the vector backbone in a modified Kunkel procedure. Using this approach, we were able to reliably generate libraries of ~10^7 cfu/ug PCR product per transformation in a single day. We have successfully employed this method to generate libraries for three different single-chain antibodies and, in conjunction with phage display, have identified variants with enhanced function from these libraries. The key advantages of this technology includes facile amplification, restriction enzyme-free library generation, and due to the single annealing step with linear amplification, a significantly reduced risk of unintended base pair changes at unintended sites within the plasmid.

2.2 INTRODUCTION

Directed evolution, the *in vitro* generation of mutants followed by function-based selection of variants, is a standard tool for many protein engineering, synthetic biology and metabolic engineering applications [130, 131]. Due to the limited ability to predict the effects of amino acid changes on protein structure and function, a variety of approaches to protein-based directed evolution have been developed, including site directed and random mutagenesis as well as recombination-based methods, such as DNA shuffling. These methods are widely used for epitope mapping of protein interactions, [112, 132, 133], analysis of protein function [113, 114], optimization of metabolic pathways through promoter engineering [134, 135], and enzyme engineering applications, including modulation of co-factor use [136, 137]. In particular, random mutagenesis is an effective tool for modulating protein activity and stability as long as a robust function-based screen is available [107, 108] and can be used to enhance nearly any biochemical or biophysical parameter, including antibody affinity for ligands [109-111].

The goal of a directed evolution experiment is to identify clones with enhanced function; the probability of success is enhanced by generation of large, productive libraries, encoding few wild-type variants, stop codons or frame shifted variants. Standard generation of libraries in *E. coli* proceeds with three steps: (1) amplification of the region of interest under error-prone conditions, (2) restriction enzyme digestion and ligation into a plasmid and (3) transformation into competent cells. Each of these steps can be time consuming and inefficient, limiting the final library size [123]. In particular, the cloning steps, involving digestion of the plasmid, amplification of insert and subsequent ligation of these two fragments, present bottlenecks during library generation. These difficulties have stimulated development of recombineering approaches which

avoid traditional cloning steps. Other techniques avoid direct DNA manipulation completely, using *E. coli* mutator strains such as DM2516 and XL1-Red (deficient in mutS, mutD and mutt; Stratagene). This method suffers from a low mutation rate (~6–10 mutations/kb/generation in our hands or a few mutations per 100 generations [125]. In addition, mutations are uniformly distributed across the plasmid, potentially impacting the promoter and antibiotic resistance gene.

Random mutagenesis procedures typically employ the polymerase chain reaction (PCR) due to its simplicity and ease of mutant generation [138-140]. In particular, Stratagene's QuikChange site-directed mutagenesis Protocol (QCM; [115]) is one of the most widely used PCR-based methods, utilizing methylated plasmid template and two complementary, synthetic oligonucleotides. The oligonucleotides both encode the desired mutations and serve as primers to synthesize the entire plasmid during a PCR-based linear amplification step. Originally developed to introduce single mutations into a gene, variations have (1) increased the plasmid PCR yield via partially overlapping oligonucleotides and exponential amplification [117], (2) introduced targeted random mutations with degenerate oligonucleotides [118], and (3) use of non-overlapping primers to introduce site-directed mutations, and generate megaprimers for PCR-amplification of the entire plasmid [119-121]. QCM has been employed for random mutations by amplifying the plasmid under error-prone conditions, although these mutations are not restricted to an area of interest [119, 122]. While the simplicity of these methods is appealing, exponential amplification of the whole plasmid increases the risk of introducing errors at essential or undesirable locations (*e.g.* origin of replication, promoter and antibiotic resistance genes) even with a high fidelity polymerase.

Alternatives to QCM, such as recombineering and ligation-independent cloning approaches avoid the difficulties of whole plasmid PCR [134, 139] but are not optimized

for generation of large libraries. Modified Kunkel mutagenesis procedures avoid exponential PCR amplification and restriction enzyme digestion/ ligation bottlenecks by utilizing template DNA containing uracils (dU-ssDNA) to form heteroduplex DNA with primers containing the desired mutations [104], which are extended to create a dTTP-containing complementary strand by T7 DNA polymerase in a linear amplification process. Upon transformation into into *dut*+*ung*$^{+}$ *E. coli*, the template dU-ssDNA is inactivated, allowing only the mutated strand to be propagated. An inexpensive and rapid technique, a key advantage to this protocol allows for simultaneous annealing of multiple independent primers to different locations on the plasmid. With targeted mutagenesis libraries, in which sequential codons are randomized with degenerate primers, libraries with over 10^{11} individual clones and eight randomized position have been constructed using this method [132].

Utilizing a combination of randomly mutated megaprimers and Kunkel mutagenesis, we propose a novel method for library generation to overcome the limitations of the current protocols. First, the region of interest is amplified under error-prone conditions. Next, an asymmetric PCR step generates a library of 3' megaprimers from the error-prone product. Last, the megaprimers are used in conjunction with dUTP-containing, single-stranded template plasmid to generate a dTTP- and megaprimer-containing complementary strand (Figure 2.1). Upon transformation, this method reliably and rapidly creates large libraries (~10^{7}/ ug DNA/ transformation) and can be scaled-up for larger libraries. We have successfully employed this method, termed MegAnneal, to generate libraries for three different single-chain antibodies (scFv) and, in conjunction with phage display, have identified variants with enhanced function from these libraries. The key advantages of this technology includes facile amplification, restriction enzyme-free library generation, and due to the single annealing step and linear amplification, a

significantly reduced risk of unintended base pair changes at unintended sites within the plasmid.

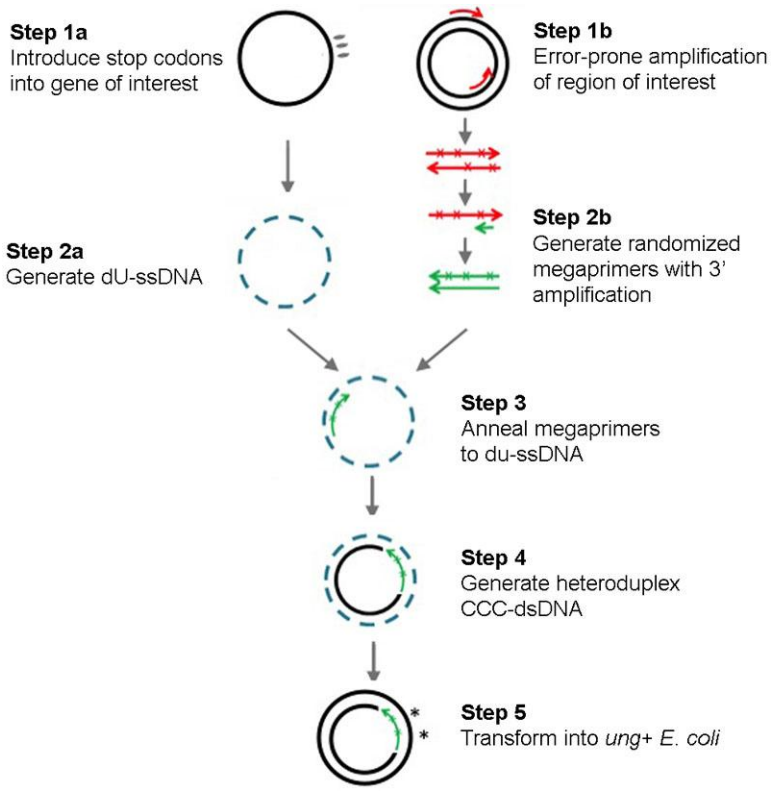


Figure 2.1 Generation of ligation-free error-prone libraries with megaprimers.

Schematic diagram of the process includes PCR mutagenic amplification of PCR in the reverse reading frame. Followed by annealing of megaprimers to uracil containing single stranded DNA (dU-ssDNA) and generation of heteroduplex double stranded DNA (ccc-dsDNA).

2.3 Materials and Methods

2.3.1 Plasmid preparation

The scFv genes (750 bp) was inserted via *Sfi*I-*Sfi*I ligation into the phagemid vector pMoPac24 (~4800bp), which contains an amipicillin resistance cassette and an

M13 origin of replication [141]. Three scFv genes were used: the 3D5/EE scFv, with initially poor expression and weak affinity for the EE peptide (sequence: EYMPME); the M2 scFv, with poor expression and affinity for the FLAG peptide (sequence: DYKDDDDK) and the hu1B7 scFv, with poor expression and moderate affinity for the pertussis toxin [142]. To minimize the presence of wild-type peptide specific scFv in randomly mutated libraries, three tandem stop codons (TAA TGA TAA) were inserted into the gene in the CDR H3 region using overlapping oligonucleotides and standard QuickChange mutagenesis. An additional three tandem stop codons (TAA TAA TAG) were inserted into the gene at CDR H2, to create the template for the combined HCDR2 and HCDR3 control library to ensure the successful annealing of both primers. Stop codons were inserted near the 3' end of the scFv, so that even short megaprimers will generate in frame scFv sequence upon annealing and extension. For the hu1B7, mutations were expected across the entire gene length, therefore stop codons were inserted at the beginning of the heavy chain of hu1B7, so that only megaprimers of full length would anneal. After confirmation of stop codon insertion by sequencing, pMoPac24 plasmid containing the scFv-stop gene was transformed into the *dut*$^{-}$*ung*$^{-}$ *E. coli* strain CJ236 (NEB). Clones were grown in liquid culture to mid-log phase at 37 C, infected with M13KO7 phage (NEB) at a multiplicity of infection (MOI) of 10 to package uracil-containing, single-stranded plasmid DNA (dU-ssDNA) within the phage particles [104]. The phage were concentrated via precipitation with ¼-volume 20% PEG-1.5 M NaCl [124] and the dU-ssDNA extracted using Qiagen's Qiaprep Spin M13 Kit for use as template DNA during the subsequent mutagenesis and library generation steps.

2.3.2 Generation of randomly mutated megaprimers

Randomly mutated megaprimers were produced via two sequential PCR steps: (1) amplification of the region of interest under error-prone conditions and (2) an asymmetric PCR step to generate 3' megaprimers. First, the entire scFv gene (10 ng) was amplified using the low fidelity Mutazyme II DNA Polymerase (Stratagene, La Jolla, CA) and two flanking primers (set 1: 5' scback and 3' scforlong; set 2: 5'pakpel and 3'hucK; see Table 3) according to commercial instructions. The 50 uL reactions were initially heated at 95^{o}C for 4min, followed by 25 cycles of incubation at 94^{o}C for 30sec, 58^{o}C for 30 sec and 72^{o}C for 1min. After amplification, PCR product (~750 bp) was purified from a 1% TAE agarose gel using QIAquick gel extraction kit (Qiagen). To calculate the mutation rate from this initial step, purified PCR products were cloned into pTopo (Invitrogen), transformed into Top10 E. coli cells and the scFv gene sequenced from 10 independent colonies to provide an initial estimate of mutation frequency.

A second PCR step was performed using only the anti-sense primer (3'scforlong) and 32 cycles to generate a library of randomly mutated anti-sense megaprimers. Megaprimers were gel purified and phosphorylated in 20uL reactions containing ~2ug megaprimer, TM Buffer (0.05M Tris, 0.01$MgCl_2$, pH 7.5), 1 mM ATP, 5 mM DTT and 5 units of T4 polynucleotide kinase incubated at 37^{o}C for 1 hour. One gene (hu1B7) was recalcitrant to a second amplification, producing only PCR products of the incorrect size. In this case, the initial Mutazyme reaction was omitted and the megaprimers produced using target DNA under biased nucleotide conditions expected to yield ~1% error rate [143, 144].

2.3.3 Library Construction

To generate error-prone libraries for directed evolution, randomly mutated megaprimers were annealed to dU-ssDNA containing the scFv-stop gene and extended in a modified Kunkel procedure. The dU-ssDNA template plasmid (20 ug) was combined with a three-fold molar excess of phosphorylated megaprimer into TM buffer for a final volume of 250uL. The mixture was heated to 90^{o}C for 2 min to disrupt secondary structure, followed by cooling at 50^{o}C for 3 min, then 20^{o}C for 5 min to facilitate annealing of megaprimer to dU-ssDNA template plasmid. Next, extension of the annealed megaprimer via T7 DNA polymerase and repair of nicked ends by T4 DNA ligase created covalent, closed circular double-stranded plasmid (ccc-DNA) comprised of a dU template strand and an anti-sense dTTP strand. The annealed DNA mixture was adjusted to contain 0.34 mM ATP, 0.85 mM dNTPs, 5.10 mM DTT, 240 U T4 DNA ligase and 3 U T7 DNA polymerase and the reaction allowed to proceed overnight at room temperature.

The resulting double stranded DNA was desalted using QIAquick gel extraction columns (Qiagen) and quantified by absorbance (1 OD_{260nm} = 50ug/mL). Successful generation of ccc-DNA was confirmed by comparison with template dU-ssDNA (500 ng each) on a 2% agarose gel. Desalted polymerase reaction (~10ug) was combined with electrocompetent XL1-Blue *E. coli* cells (500uL, 10^{8} cfu/ug competency), electroporated in a 2-mm gap electrocuvette (Biorad, Hercules, CA) at 2.5kV, 25μF, 200Ω and immediately recovered in 2mL of TB media at 37^{o}C for 45 minutes with shaking before transfer to 500mL TB with 200ug/mL ampicillin. To determine library size, aliquots were plated on selective agar plates. Three scFv libraries using the 3D5/EE gene were produced using megaprimers of varying lengths (250bp, 750bp and a pool >150 bp) to focus mutagenesis to different regions of the scFv gene as well as determine the effect of

megaprimer length on final library size and diversity. In addition, two control libraries were created using either one or two 50 bp oligonucleotides in place of megaprimers (see Table 3). These oligonucleotides introduce 12 and 7 randomized codons in the CDR H2 and H3 regions, respectively, and provide a calibration point for library size with respect to standard Kunkel mutagenesis.

2.3.4 Analysis of library diversity and quality

After library generation, 30-50 clones per library were sequenced with primer 5'pakpel to measure the expected scFv mutation rate, percent of clones incorporating megaprimer, and rates of mis-priming and frameshifting. To assess the rate of unintended mutations introduced by T7 polymerase-mediated linear amplification of the entire plasmid, a region of the backbone plasmid (skp region using primer 5'skp) was sequenced from 10 clones.

Following library generation, production and panning of recombinant scFv-displaying M13 phage was employed to identify clones with enhanced solubility and EE-peptide binding affinity, as well as demonstrate the quality and richness of the megaprimer-based library. After transformation, the pooled library was grown with shaking at 37^{o}C to mid-log phase. ScFv expression was induced with 1mM IPTG and the temperature shifted to 25^{o}C for 3 hours. Next, M13K07 helper phage were added (MOI of 10) and allowed to rest for 30 minutes, followed by two hours of growth and addition of kanamycin (25ug/mL) prior to overnight growth to produce scFv-displaying phage particles (~12 hours). Phage were purified by double-precipitation with a ¼ volume of 20% PEG, 2.5 M NaCl, resuspension in PBS and quantification by absorbance at 260 nm.

Phage panning was performed in ELISA (Costar) wells coated with 4 ug/ml anti-c-myc antibody (9E10, Sigma) or target ligand (maltose binding protein, MBP, with a terminal EE tag) and blocked with 5% non-fat milk in PBS. Panning consisted of three selection rounds: one with immobilized anti-c-myc antibody to enrich for full-length scFv and remove variants with stop codons or frameshifts, followed by two rounds with peptide-tagged ligands or pertussis toxin. For each round, 10^{12} plaque forming units (pfu) were added to a coated, blocked ELISA well. After equilibration for 1 hour at 37^{o}C, and washing 10 times with PBS-0.05% Tween, bound phage were eluted with 0.1 M glycine-HCl pH 2.2, transferred to a new tube and neutralized with 2 M Tris, pH ~10.7. Next, 2 ml of E. coli (ER2738) were infected with the eluted phage to amplify selected clones. A fraction of this volume was plated to estimate the number of pfu recovered and to isolate single clones.

To characterize individual clones selected during panning, single colonies were inoculated into 150uL of TB with ampicillin (200ug/mL) in a sterile 96 well plate and phage produced as above. Plates were centrifuged and 50uL of phage-containing supernatant used in activity assays. Two ELISA assays were used to characterize clones selected during panning. First, incorporation of full-length scFv protein into phage particles was monitored via binding of a c-terminally encoded c-myc peptide tag to immobilized anti-c-myc antibody (4 ug/mL). Next, phage presenting active scFv was captured by immobilized MBP-EE ligand (10 ug/ml). In each assay, bound phage were detected using anti-M13-HRP (1:5000, GE Healthcare) and signal developed with TMB substrate, quenched with 1N HCl for absorbance measurements at 450 nm. Since phage displayed scFv variants can be selected based on enhanced expression as well as affinity, the specific activity (the ratio of MBP-EE to anti-c-myc binding activity) was used to normalize for expression differences and facilitate comparison between clones. Negative

controls included uncoated, blocked ELISA wells and plasmid containing the scFv-stop gene, while positive controls included scFvs known to present well on phage.

2.4 RESULTS

2.4.1 Error-prone scFv gene amplification

As a first test of megaprimer-based generation of random mutagenesis libraries, we started with the 3D5/EE scFv, aiming to identify variants with improved EE-peptide affinity, expression and solubility. Our ultimate goal is that well-behaved scFvs with EE peptide binding specificity could be used as crystallization chaperones for hard-to-crystallize proteins presenting with EE peptide. In principle, any PCR-based method for introducing errors into a gene, such as biased nucleotides and error-prone polymerases, would be compatible with the megaprimer method. We initially selected the error-prone Mutazyme thermophilic polymerase, as yields of amplified, randomly mutated DNA were ~10-fold higher than with other methods. In addition, the mutagenesis rate, based on amount of template DNA in the PCR reaction, was easier to control than with biased nucleotides or plasmid amplification in mutS deficient XL1-Red, which yielded few mutations after 15 rounds of cell growth.

Amplification of the scFv gene with either set 1 or set 2 primers and mutazyme polymerase produced a sharp band of the expected size (~750 bp) and ~2ug total DNA. After gel purification, the DNA was ligated into pTopoTA (Invitrogen), transformed into Top10 cells and 10 clones sequenced with M13 Forward primer. Using the error-prone Mutazyme polymerase, mutation rates are controlled by the quantity of plasmid introduced into the reaction, with 10-100ng expected to result in 3-4 mutations per 1000bp (Stratagene). Up to 10 mutations were observed in select colonies, with an

average of three mutations per scFv gene. While the mutations were present in both the heavy and light chains of the scFv, the mutations were more frequently observed in the heavy chain.

2.4.2 Generation of error-prone megaprimers

Purified product from the mutagenic PCR reaction (~400 ng) was used as template DNA in a second, asymmetric PCR reaction with only the 3' scforlong primer and Vent polymerase. In the presence of a single primer, PCR linearly amplifies the template DNA during each cycle. Under these conditions, incorrect primer annealing or enzyme processivity can result in premature termination of the growing DNA strand; as a result the PCR product can exhibit a range of sizes. Utilizing the same 3' huCk primer as used in the mutagenic reaction produced poor yields with multiple products, while use of a nested primer, 3' scfor, which anneals 46 base pairs internal to the original 3' huCk, resulted in strong bands of a single size corresponding to the full-length scFv gene (Figure 2.2). In the case of hu1B7 gene, products of the expected size were never obtained upon amplification of Mutazyme product. In this case, we used an alternative approach: direct production of error-prone megaprimers in a 3' asymmetric PCR reaction performed with biased nucleotides to introduce random nucleotide base pair changes.

Products generated from this second reaction were separated on an agarose gel, excised and purified to generate megaprimers for use in the library generation step. PCR products using the nested set 2 primers (Figure 2.2) produced a single band at 750 which was excised and purified. For PCR products using the non-nested set 1 primers, which resulted in a range of product sizes (Figure 2.2), the megaprimers were separated into three different primer sets of varying lengths: i) 250 only, ii) 750 only, iii) a pool >150

bp. Each discrete megaprimer size was used to generate a different library, as indicated in Table 2.1. For hu1B7, in which error-prone megaprimers were produced in a single PCR step, megaprimer size was controlled by extension time and produced discrete bands at ~750 and 1500 bp, which were combined to produce library VI.

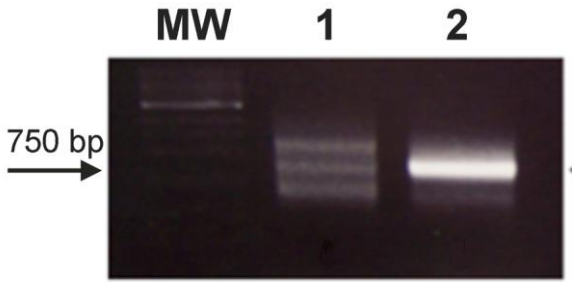


Figure 2.2 Megaprimer production.

PCR product produced by asymmetric amplification of randomly mutated scFv genes with the primer 3' scfor. Yield and quality of the megaprimers varies with the annealing position of the 3' primer used: lane 1, identical primer used in the initial error-prone amplification reaction and lane 2, use of a nested primer.

2.4.3 Generation of megaprimer-containing ccc-DNA

Five separate libraries were constructed to examine the effect of megaprimer length on the resulting library size and diversity. Three libraries used megaprimers with different lengths: (i) 250bp; (ii) 750bp; (iii) mixture of lengths over 150 bp. Two libraries were control libraries using (iv) a single 50bp oligonucleotide (primer 3' HCDR3) and v) two 50 bp primers used simultaneously (3' HCDR2 and 3' HCDR3), constructed via standard Kunkel mutagenesis. The final library was constructed using biased nucleotide mutagenesis to compare 3'megaprimer generation methods and the impact on size and diversity. The megaprimers were annealed to dU-ssDNA plasmid containing the gene of interest with three tandem stop codons near the 3' end of the gene. Next, reactions with

T7 DNA polymerase extended the megaprimer to generate a heteroduplex, consisting of a dUTP-containing strand and a complementary dTTP-containing strand. Finally, T4 DNA ligase repaired the nicked ends, prior to desalting and transformation into competent cells. Once within *dut*+ *ung*+ bacteria, the dUTP-containing DNA strand is degraded and only the dTTP- and megaprimer-containing stand retained and propagated as a plasmid. Each megaprimer preparation was able to prime T7 polymerase for whole plasmid amplification, as monitored by agarose gel analysis of the product (Figure 2.3). The ccc-DNA runs as three bands (open-circular, supercoiled and nicked) which migrate more slowly than the dU-ssDNA. The yield and quality of the ccc-DNA were predictive of final library size and quality and serves as check during library development.

Table 2.1 Library size based on megaprimer number and length.

Library	# primers	Megaprimer length (bp)	Cfu/ ug DNA/ transformation	% clones encoding full-length scFv	scFv error rate (%)*
I	1	250	1.0×10^7	79.2	1-2bp (0.1-0.2%)
II	1	750	9.3×10^6	41.3	4-7bp (0.5-0.9%)
III	1	>150	1.1×10^6	54.9	1-4bp (0.1-0.5%)
IV	1	50	4.1×10^7	50.0	-
V	2	50	1.1×10^7	37.0	-
VI	1	750 and 1500	10^6	62.5	2-5 (0.2-0.67%)

*range observed based on 30-50 colonies sequenced per library

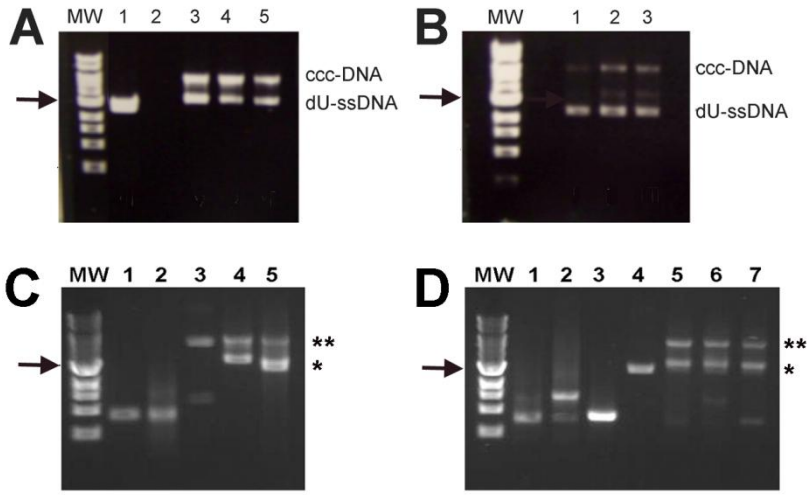


Figure 2.3 Generation of ccc-DNA by megaprimers.

(a) Control libraries using 50 bp primers and randomizing 7 or 12 codons within scFv 3D5/EE. Lane 1: dU-ssDNA with stop codons in the HCDR2 region, lane 2: blank, lane 3: ccc-DNA product from primer HCDR2 (50bp), lane 4: ccc-DNA product with primer HCDR3 (50bp); and lane 5: ccc-DNA production with HCDR2 and HCDR3.
(b) Effect of the megaprimer length on ccc-DNA production for scFv 3D5/EE. Lane 1: 750bp; lane 2, >150 bp and lane 3: 250bp megaprimers.
(c) Megaprimer and ccc-DNA production for scFv M2. Lane 1: mutazyme reaction product with set 1 primers; lane 2: megaprimer production with nested 3' primer; lane 3: dU-ssDNA, lane 4: ccc-DNA with site-directed oligonucleotides; lane 5: ccc-DNA produced with error-prone megaprimers.
(d) Single-step production of error-prone ccc-DNA for scFv hu1B7. Lane 1: megaprimer produced with Taq and biased nucleotides (700 bp); lane 2: megaprimer produced with Taq and biased nucleotides (1500 bp); lane 3: megaprimer produced with Vent and biased nucleotides (700 bp); lane 4: hu1B7 dU-ssDNA; lane 5: ccc-DNA produced with 700 bp Taq megaprimers; lane 6: ccc-DNA produced with 1500 bp Taq megaprimer; lane 7: ccc-DNA produced with 700 bp Vent megaprimers. Arrow indicates the 3000 bp molecular weight marker, * dU-ssDNA and ** ccc-DNA expected sizes.

2.4.4 Rapid generation of large, rich error-prone libraries without restriction enzymes

After confirmation of successful generation of ccc-DNA production, DNA (~10 ug) was transformed into electrocompetent *E. coli* cells with a single transformation (100

uL of cells with a competency of 10^8 cfu/ug/uL) to generate a library. Aliquots of the transformation were plated to estimate library size, while the remainder was use to produce recombinant scFv-displaying phage for panning experiments. In general, a single transformation produces $>10^7$ transformants/ ug DNA (0.2 cfu/ ug PCR product/ cell), which can be readily scaled up to produce very large libraries. Sequencing confirmed the presence of a full-length scFv in about half the transformants (range, 37-79%), consistent with standard Kunkel results and a similar error rate within the scFv gene as expected from sequencing of the Mutazyme reaction (0.1-0.9%; Table 1). Increasing the megaprimer:DNA ratio from 3:1 to 5:1 doubled the final library size for library II, but further increases did not yield any additional gains.

Library size was impacted by megaprimer length using this method, with shorter megaprimers being more efficiently incorporated (Table 1). Libraries generated with megaprimers smaller than 300bp (I, IV and V) produced libraries with 10-fold higher cfus than those using larger megaprimers (II, III, VI). Analysis of the quality of our ccc-DNA derived from each megaprimer (Figure 2.3b) shows stronger and more distinct product bands for smaller megaprimer. This suggests that shorter primers anneal more efficiently to the dU-ssDNA, while larger megaprimers take longer to find the cognate sequence and may partially anneal. The variable length megaprimer library (III) is about 10-fold smaller than the others, which may result from competition between primers of different length for template plasmid, with shorter primers annealing faster to the ssDNA. Sequences from library III revealed mutations primarily in the heavy chain region (~430 bp region), supporting the idea that short megaprimers outcompete longer ones.

To determine the actual library mutation rate and assess whether the Kunkel mutagenesis step introduces additional mutations within the plasmid, 30-50 clones per library were sequenced at two separate locations. Sequencing was performed with primer

5’ pAKpEL which sits upstream of the scFv gene and with primer 5’ skp, which sits in a region of the plasmid distal from that targeted for mutation (Table 2.1). Sequencing of 30 clones from Library II with 5-pakpel primer revealed an average error rate of 0.73% (4-7 base pair changes per gene), with mutations distributed across the heavy and light scFv chains and expected levels of transitions and transversions as described by manufacturer. Final library mutations will be constrained to the region of the gene covered by the megaprimer; thus while 250 and 750 bp megaprimers may each encode 1% errors, the shorter megaprimer will confer a lower overall mutation rate to the gene since only one-third of the scFv gene is targeted. Taking this into account, the number mutations observed within megaprimer-encoded regions are statistically similar to that observed after the mutazyme reaction. Furthermore, no mutations were observed in the skp gene, a region of the plasmid not targeted for mutagenesis. Megaprimer mispriming was observed exclusively in library III, in which several colonies (4 of 51 sequenced), encoded reading frame shifts and one non-target gene insertion. This may be a result of the size variation of these megaprimers, which also constrained library size or mis-priming during megaprimer generation.

For generation of optimal libraries, megaprimers must be clear strong bands with DNA only in the 3’ direction. The intended megaprimers should not include PCR product that is the result of primer misannealing (i.e. smears, indistinct bands). Additionally the megaprimers should not be larger than the expected size of your gene, as this leads to mis-priming. From our data, it is suggested that megaprimers of 250bp and less are more effective at generating large libraries (10-fold higher) than those with similar quality megaprimers at a length of 750bp. Should multiple sized megaprimers bands occur that fall within the range of your gene length, the recommendation is to anneal each megaprimer size independently to avoid competition issues between the

range of primers. After transformation, these colonies can be combined into one library before proceeding to the selection/screening phase.

After two rounds phage panning and selection, pooled clones from library i-iii were screened for scFv expression and EE-peptide binding ability. Selected colonies were grown in 96 well plates, infected with M13 phage to produce scFv-displaying phage and the phage used in two ELISA binding assays. Binding to anti-c-myc antibody indicates the presence of a full-length scFv protein incorporated into the phage coat, as the c-myc epitope is encoded at the c-terminus of the scFv. Binding to the target EE peptide reflects the scFv activity, while the c-myc/ EE ratio provides a measure of specific binding activity. Of the 96 colonies screened, 15 bound anti-c-myc and showed successful expression of full length and completely folded proteins (Figure 2.4). Of these 15 clones, 11 exhibited varying degrees of affinity for the EE peptide (Figure 2.4). In contrast, the WT gene (A12) produced low-level signal for binding to anti-c-myc and was undetectable on the EE screen. Notably, clone A2 exhibited remarkably high specific activity (c-myc: EE ratio of 6), indicating the presence of high-quality variants in the mega-primer-produced library. Further analysis of soluble protein a variants selected from this library characterized variants with 38-fold higher affinity, six-fold greater solubility (2.3 vs 12.8 mg/ml) and enhanced monomeric expression levels (50% vs 81%; Pai et al, submitted, Chapter 3).

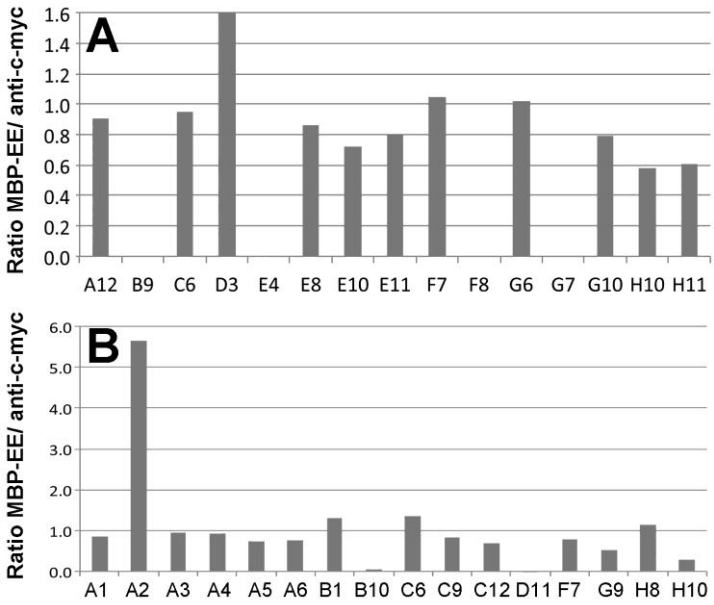


Figure 2.4 Functional analysis of clones selected from 3D5/EE MegAnneal libraries.

Individual clones were grown and recombinant scFv-displaying phage produced in 96-well plates, prior to use of supernatant in phage ELISAs. Shown is the specific activity of each clone, calculated as the ratio of absorbances from EE and anti-c-myc ELISAs. (a) library I (b) library II. Controls: WT 3D5/EE (clone A12), plasmid with scFv-stop gene (clone B3).

2.5 DISCUSSION

Overview Kunkel mutagenesis was developed as an early method to introduce site-directed mutations in genes [145] and has been used to produce site-directed libraries as large as 10^{11} independent clones randomizing eight sequential amino acid residues [132]. We have adapted Kunkel mutagenesis to generate error-prone libraries for directed evolution of select regions or an entire gene (Figure 2.1). First, template plasmid is prepared by introducing three stop codons at the 3' end of the original gene using point mutagenesis to prevent wild-type contamination of the final library (step 1a). This template is then transformed into CJ236 *ung*$^-$ *dut*$^-$ *E. coli* cells followed by M13 phage

super-infection to produce dU-ssDNA template plasmid containing an inactive scFv gene (step 2a). Simultaneously, random mutations are incorporated into the gene-of-interest by exponential amplification with a low-fidelity polymerase, biased nucleotides, or other method (step 1b). The resulting PCR product is purified and linearly amplified with a nested 3' primer to enrich for antisense strands capable of annealing to the dU-ssDNA (step 2b). Megaprimers anneal to the template DNA, priming T7 polymerase to synthesize a dTTP- and megaprimer- containing copy of the template plasmid, with the ends covalently joined by T4 DNA ligase (step 3). While we show use of a single megaprimer here, two could be used to simultaneously randomize multiple regions of the gene or plasmid. Complementary base pairing between template and copied strands results in covalent-closed-circular double-stranded DNA (ccc-dsDNA), whose presence can be readily monitored by gel electrophoresis (step 4). Upon transformation into *ung*$^+$ *E. coli* cells, such as XL1-Blue, the uracil containing strand is efficiently restricted by native nucleases while the mutant plasmid strand is replicated and propagated (step 5), providing a second level of background template elimination.

2.5.1 Megaprimer amplification versus competing technologies

MegAnneal allows for rapid production of large and diverse error-prone libraries with mutagenesis targeted to one or more locations on a plasmid and covering regions 150 to 1500 bp in size. Constraints of the method include the requirement for an M13 origin of replication on the plasmid, introduction of stop codons within the template DNA and production of dU-ssDNA template plasmid. These steps require three days of advance work, but are crucial for minimizing the wild-type contamination of the library. With dU-ssDNA template in hand, we are able to rapidly create large libraries (10^7 cfu

/ug ccc-DNA with 50uL of 10^8 electrocompetent cells or 0.2 cfu/ ug ccc-DNA/ competent cell) within 24 hours. Importantly, the ccc-DNA can be stored indefinitely, and simply re-transformed to produce a fresh library for screening, reducing the effect of growth differences among clones upon repeated growth of frozen libraries.

A comparison of MegAnneal and competing methods such as QuikChange, Megawhop and standard Kunkel Method (Table 2.2) underscore the advantages of this technique in terms of (1) uniform error via megaprimer production with a low fidelity polymerase or error prone conditions; (2) minimized mispriming as a result of the slow megaprimer annealing process; (3) reduction in unintended mutations outside the region of interest with the use of linear amplification and (4) library contamination by wild-type genes is limited by the use of two orthogonal mechanisms. Importantly, the limitations of Kunkel mutagenesis also apply to MegAnneal. The vector chosen must be compatible with not only the method of screening but also the method of dU-ssDNA generation, and thus require an M13 origin of replication in order to package dU-ssDNA. Since strain CJ236, used to produce du-ssDNA, encodes chlorampenicol resistance, this should not be the plasmid selectable marker as well.

Table 2.2 Comparison of different mutagenesis techniques used for library generation.

Technique	# mutagenic primers	Primer length Range (bp)	Controllable mut freq	Codon-based mutagenesis	Limitations	Ref.
QuikChange	1	~50	Yes	Yes	Primer quality, Requires PCR Optimization	[115]
MEGAWHOP	1	500-3000	Possible	No	Numerous Primers & Amplification Steps Requires PCR Optimization for each primer Steps	[119]
Kunkel Method	unlimited	~50	Yes	Yes	Choice of Vector	[104, 145]
REM with ligation	0	N/A	No	No	Clean digestion of vector and insert	[130]
Shuffling	NR	NR	No	No	Homologous Genes required	[146, 147]
MegAnneal	**multiple**	**50 – 1500**	**Yes**	**Yes**	**Choice of Vector**	**-**

While error-prone amplification with Mutazyme was the primary method for random mutagenesis used here, any ranodom mutagenesis procedure is compatible with MegAnneal. Mutazyme II DNA polymerase was selected for the insertion of randomized mutations due to its ability to produce more uniform mutations without apparent bias for any particular nucleotide or amino acid which is commonly seen in other error-prone PCR methods, unlike Taq polymerase (which restricts mutations to nucleotides A and T) or weighted concentrations of dNTPs to introduce randomization [119, 138]. However, random mutagenesis is inherently limited by the unequal distribution of codons to amino acids and their natural frequency in protein sequences [130]. Mutazyme's ability to construct low to medium rate mutation libraries dependant on the initial amount of target

DNA make it an extremely powerful tool to compare with other error-prone PCR protocols [108]. By varying the initial concentration of DNA template in the reaction (0.1-1000ng), a large range of mutation frequencies (1-16 mutations/kb) can be achieved allowing for nearly complete control over the mutation rate per gene while maintaining high PCR product yields with varying fold amplification (1-10,000x). While this particular polymerase is more costly than standard PCR polymerases, the reaction conditions have been optimized by Stratagene and only small (50 ul) reactions are required.

We directly compared the reaction products from QuikChange and MegAnneal (Figure 2.5). There are two major disadvantages to Quikchange: primer dimer formation and mispriming during whole plasmid amplification. Both of these problems are a function of the primer design and length. Primer dimer formation occurs frequently in Quikchange and kinetically competes with the desired product during primer annealing. In MegAnneal, the use of a single primer initiating megaprimer amplification largely eliminates the risk of intermolecular dimer formation. The coincidental match in DNA sequences at these primer lengths (250-950bp) is relatively small and chances of mispriming are unlikely. The slow cooling during annealing allows for megaprimer-template hybridization; nicontrast, rapid temperature changes potentially cause non-specific and mismatched hybrids [148]. Not only does it benefit hybridrization and reduced mispriming, it reduces the need to optimize the T_a value for this reaction which is why multiple primers can be annealed simultaneously. Finally, the single annealing and amplification cycle prevents an inexact pairing of megaprimer and dU-ssDNA to be amplified in subsequent cycles, which occurs with multiple cycle PCR-based methods.

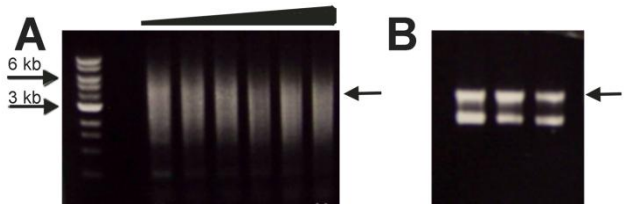


Figure 2.5 Quick change vs meganneal for targeted mutagenesis.

(a) (standard quick change eg for stop codons insertion) PCR Product using QuickChange method with overlapped oligos at varying annealing temperatures from 47 to 52^{o}C.

(b) Our polymerase product after control primers (50bp) were annealed to dU-ssDNA and complementary synthesis of Lane 1: ccc-DNA product from primer HCDR2 (50bp), lane 2: ccc-DNA product with primer HCDR3 (50bp); and lane 3: ccc-DNA production with HCDR2 and HCDR3, desired band is indicated.

The linear whole plasmid amplification provides additional benefits due to the reduced risk of unintended mutations at extraneous sites. For exponential PCR-based methods, amplified copies of plasmid are used as template in subsequent steps, thus any error will be propagated by daughter copies. Unintended mutations, for instance in the origin of replication, the promoter or the antibiotic resistance gene, may result in loss of a clone from the library or diminished ability to compete due to reduced protein expression. The amplification in this protocol has even less relative risk of DNA base pair changes outside the intended region because T7 polymerase error rate is fairly low (error rate: 1.5 x 10^{-5} per base incorporated, NEB) and each template plasmid is amplified at most once. This results in easy quality control during library construction, allowing the investigator to readily monitor the quantity and quality of CCC DNA. In comparison, thermophillic polymerases used in exponential whole-plasmid amplification protocols such as Taq (error rate: 2.28×10^{-5} per base, NEB) or even high fidelity Pfu (error rate: 4.4×10^{-7} per base, NEB) can introduce mutations that are propagated exponentially with each additional cycle.

Many site-directed and random mutagenesis only have one mechanism to limit contaminating template genes in the library. Many methods, including QuickChange, use methylated template DNA produced from *dam*$^{+}$ *E. coli* strains or in vitro methylation with CpG *Sss*I enzyme. In either case, methylated template DNA is digested *in vitro* with methyl-specific *Dpn*I restriction enzyme (target sequence: 5'G^{m6}ATC-3') prior to transformation (QuikChange). In our hands, this results in recovery of 25% wild-type template. Our process uses a comparable method with uracil containing dU-ssDNA which is inactivated by *E. coli ung*$^{+}$ strains, where literature has shown that similar amounts of wildtype template or eliminated, reported here as 20-50% template recovered from a library. However, the additional step of inserting stop codons into our template plasmid eliminates the risk of wildtype phenotype controlling the library.

2.5.2 Megaprimer considerations

Our results demonstrate that megaprimers are a key component to this method with DNA fragments up to 750bp successfully annealed. This particular parameter can be optimized to impact the size of the resulting library: i) primer design –at both the randomization step and the megaprimer; ii) megaprimer length and number iii) megaprimer to template dU-ssDNA plasmid. Significant consideration needs to be taken into account when designing the flanking primers used during randomization in addition to the antisense primers required to produce the 3' megaprimers. When designing the 3' end primers it is important that the 3' primer used during the random mutagenesis sets down further away from the gene than the 3' primer used during the anti-sense amplification. As we mentioned previously, the use of the identical primers in both steps produces the DNA fragment of interest in addition to other PCR products (Figure 2).

Using a distinct second 3' primer during the anti-sense amplification that sits slightly further down on the randomized fragments, better annealing is achieved and higher quantities of the intended DNA fragments are produced. Limiting the amount of complementary 5'megaprimer present during the annealing step will facilitate the successful annealing of the megaprimer to the ss-DNA, as DNA kinetics tend to favor smaller DNA fragments.

Megaprimer length can largely impact the library size that can be generated. From our three random libraries (I-III), using megaprimers <250bp produced a library (~10^7 cfu) that successfully annealed twice as well as those utilizing larger megaprimers. Libraries utilizing single megaprimers <100bp produced large libraries (~10^8 cfu when used with 10^9 electrocompetent cells). The correlation between the length of the randomly amplified insert and the transformation yield is most likely a result of the kinetic ability for smaller oligonucleotides to anneal more quickly and efficiently. This is firmly supported by the results from both library ii and iii indicating that the library with mixed lengths of primer yielded a higher percentage of successfully annealed clones (55% versus 41%). Although the maximum primer length used here was 750bp, other literature has indicated the effective use of megaprimers as large as 1.3kb (EZClone, [119]) for annealing to plasmids. Using oligonucleotides of 50bp in length, showed no notable impact on transformation yields when annealed independently or simultaneously. Careful consideration needs to be taken when determining regions of randomization so that nucleotide sequences are not similar enough to induce non-specific annealing (i.e. high GC content, hairpins, repetitive sequences, etc). This non-specific annealing was prevalent during our library generation when truncated scFvs emerged with Vh domains only.

The ratio of phosphorylated oligonucleotide to dU-ssDNA present in the annealing reaction plays a critical role in the resulting library size. From our results, we have shown that increasing the original ratio from 3:1 to 6:1 (oligonucleotide:template) doubled the library size unless the original 3:1 ratio had reached the maximum library transformant size of 10^7 cfu. It was shown that attempts to increase the ratio up to 9:1 actually reduced the number of transformants produced. Given the significant single stranded DNA present following the complementary synthesis reaction (Figure 4) the megaprimer to template ratio can still be improved. Like previous literature [116], we believe that this ratio must be optimized in the range we have studied to produce even larger libraries than those presented here.

2.6 CONCLUSIONS

Directed evolution experiments follows one of two paths: site directed mutagenesis and random mutagenesis. Site-directed mutagenesis is historically more labor intensive, but modern high-throughput screening technologies are facilitating rapid generation and analysis of large numbers of altered genes. Here, we present a novel restriction enzyme-free method for generating large diverse libraries with random mutations distributed across all or part of a gene of interest. Use of an in vitro generated megaprimer allows us to easily and inexpensively create large primers, while linear amplification of a single cycle to synthesize complementary DNA also allows reduced unintended mutations in regions outside of our intended target. Meganneal has a wide range of applications that extend to DNA shuffling and simultaneous mutations at multiple sites.

Acknowledgements The author thanks Jeong-min Hyun and Carolyne Smith for her assistance in analyzing library sequences.

Table 2.3 Oligonucleotides used in this study

Primer name	Use	Primer sequence (5'→3')
5'pAKpel	sequencing	5'-ATGAAATACCTATTGCCTAC-3'
5'sc back	sequencing	5'-GCATTACTCGCGGCCCAGCCGGCCATGGCG-3'
5' 3STP_3D5_ CDR3for	Stop codon insertion into HCDR3	5'- TCAAGGAACCACAGTCACCGTCTCGGCCTCG -3'
3'3STP_3D5 _CDR3rev	Stop codon insertion into HCDR3	5'- GGTTCCTTGACCCCATTAAGCTCAGCTTTAACTTTCACAGTA GTATA CG -3'
5'3STP_3D5 _CDR2for	Stop codon insertion in HCDR2	5'- AGGGCCACACTCACCGTGGATAAGTCCAGCAGC -3'
3'3STP_3D5 _CDR2rev	Stop codon insertion in HCDR2	5'- GTGTGGCCCTGCCCTAGAACTTTTAGTTTTAGCTCGTGCCG CCATT ATTCG-3'
3' Heavy CDR2 library	library generation	5'- CACGGTGAGTGTGGCCCTSNNCTTVNHMNYSBNGTTATASN NSNNSNNSNNAYYSNNARDMYDAATSNNTCCGATCCACTCC AGACC-3'
3' Heavy CDR3 library	library generation	5'- CCTTGACCCCAGTAATCCATAGCSNNSNNSNNGCTSNNSNN SNNSNNABNTGCACAGTAGTATACG-3'
3'hucK	Megaprimer, sequencing	5'-GGCGGGAAGATGAAGACAGATG-3'
3'scfor	Megaprimers	5'-GTATTAGGAATTCGGCCCCCGAGGCC-3'

CHAPTER THREE: Conversion of scFv Peptide-binding Specificity for Crystal Chaperone Development

3.1 CHAPTER SUMMARY

In spite of advances in protein expression and purification over the last decade, many proteins remain recalcitrant to structure determination by X-ray crystallography. One emerging tactic to obtain high-quality protein crystals for structure determination, particularly in the case of membrane proteins, involves co-crystallization with a protein-specific antibody fragment. Here, we report the development of new recombinant single chain antibody fragments (scFv) capable of binding a specific epitope that can be introduced into internal loops of client proteins. The previously crystallized hexa-histidine specific 3D5 scFv antibody was modified in the complementary determining region and by random mutagenesis, in conjunction with phage display, to yield scFvs with new biochemical characteristics and binding specificity. Selected variants include those specific for the hexa-histidine peptide with increased expression, solubility (up to 16.6 mg/ml) and micromolar affinity, and those with new specificity for the EE hexa-peptide (EYMPME) and nanomolar affinity. Complexes of one such chaperone with model proteins harboring either an internal or terminal EE tag were isolated by gel filtration. The 3.1 Å resolution structure of this chaperone reveals a binding surface complementary to the EE peptide and a ~52 Å channel in the crystal lattice. Notably, in spite of 85% sequence identity, and nearly identical crystallization conditions, the engineered scFv crystallizes in a different space group than the parent 3D5 scFv, and utilizes two new crystal contacts. These engineered scFvs represent a new class of chaperones that may eliminate the need for *de novo* identification of candidate chaperones from large antibody libraries.

3.2 INTRODUCTION

Even though the numbers of protein databank entries continue to increase, numerous proteins are rejected from the pipeline leading to structure determination. Specifically, there is a need for strategies to overcome the crystallization limitation, especially for membrane proteins and proteins with inherent conformational variability. A number of strategies to improve the likelihood of growing crystals of so-called "difficult" proteins have emerged over the last decade. Beyond improvements in recombinant expression and protein purification that enable more expansive crystallization trials, these techniques either involve modifying the protein to be crystallized in a way that improves its properties for crystallization, or introducing a second protein, a crystallization chaperone, to provide the crystal lattice. The former category includes random mutagenesis and homolog shuffling [78, 149, 150], limited proteolysis to generate a compact, stable protein entity [151], the identification of ligands to optimally stabilize a particular conformation of the protein [47], modification of the protein surface to reduce entropy [67, 68, 70], and protein symmetrization by crosslinking [75], among others.

The chaperone category involves the formation of a specific complex between a client protein and a soluble protein that provides hydrophilic residues to form crystal contacts and thus increases the chances of growing well-ordered, highly diffracting crystals of the complex. Since the first report of a crystallization chaperone used to determine the HIV capsid protein structure [81], efforts have focused on generating complexes between membrane proteins, which suffer from particularly unfavorable surface properties for crystal formation. Noncovalent complexes of target membrane

proteins with tailored antibody fragments [7, 86, 88, 89, 94, 98, 152], affibodies [80], VHH camelid domains [85] and designed ankyrin repeat proteins (DARPins, [91, 92, 153]) have been reported. In general, crystallization chaperones recognize native membrane protein sequence, and require identification of a new chaperone for each protein of interest. Fusion to or insertion of a chaperone into a flexible loop has also been described [61, 93, 94, 152, 154, 155]. The location of the fusion protein is key, as long linkers confer flexibility typically detrimental to crystallization [154]. Ideally, the chaperone should not interfere with activity or function of the client protein of interest. Nevertheless, in principle, any stable soluble protein tethered to or with high affinity for the membrane protein of interest could be used in cocrystallization experiments.

Here, we describe the first steps in development of a generalizable approach to chaperoning crystal growth: antibody fragments that can be used as a co-crystallization chaperone for any protein in which a short peptide sequence, the EYMPME epitope (EE), is inserted. We selected the hexa-histidine (His_6)-specific 3D5 scFv as the framework for protein engineering because it does not employ complementary determining region (CDR) residues in major crystal contacts and the CDRs face a wide channel that could accommodate a client protein [99-101]. We hypothesized that these CDRs could be modified to recognize a new peptide epitope, and a peptide or client protein could bind, without compromising existing crystal contacts. Nevertheless, 3D5 possesses several short-comings that require optimization for use as a crystallization chaperone including low affinity (Kd ~1 μM) for only extreme *c*-terminal histidines [99], pH sensitive binding [101], relatively poor expression in *E. coli*, and limited solubility [99]. Lastly, terminal His_6 tags, which are commonly used for protein purification, are not always removed before crystallization and can degrade over time due to low-level protease contamination. These features limit the broader application of the His_6 tag as a receptor for a

crystallization chaperone and motivate conversion of 3D5 to new peptide specificity. The EE peptide (EYMPME) was chosen for (1) its short length, (2) the presence of tyrosines to form hydrophobic interactions and hydrogen bonds that commonly dominate protein-protein binding energetics [156], (3) charged residues to form electrostatic interactions, (4) the presence of a proline to restrict conformational diversity, and (5) the availability of high affinity commercial antibodies binding these peptides [106, 157] (Covance, Sigma). Indeed, our optimized scFv exhibits enhanced crystallization propensity, including elevated solubility, stability, affinity and the ability to bind internal peptide sequences. This engineered, peptide-binding scFv represents a new class of crystallization chaperones that may eliminate the need for *de novo* identification of candidate chaperones from large antibody libraries.

3.3 Material and Methods

3.3.1 Molecular biology and expression of proteins presenting EE and His_6 peptides

Antibody binding sites (peptide sequences) were incorporated into proteins of interest via site-directed mutagenesis with mutagenic oligonucleotides (Integrated DNA Technologies). To generate a ligand with a *c*-terminal EE tag, maltose binding protein (MBP) was amplified from the *E. coli* genome, appended with an EE tag and cloned into the pAK400 vector [158], with or without a stop codon before the vector encoded His_6 tag to generate the MBP-EE and MBP-EE-His_6 ligands. The variant without a His_6 tag was used during phage screening and panning, both variants as well as a third in which MBP was cloned directly into pAK400 (MBP-His_6), were used in Western analysis. To account for steric accessibility, we generated ligands with varying numbers of internal EE tags by introducing tandem repeats of the EE coding sequence into the flexible linker

connecting the heavy and light chains of unrelated scFv proteins. A single repeat was introduced into the DO11.10 scFv gene to generate scFv-EE_1, two repeats into the 14B7 scFv gene to generate scFv-EE_2 [39] and three repeats into a non-native scFv consisting of the 3D5 light chain and 14B7 heavy chain to generate scFv-EE_3. These proteins also contain c-terminal, vector encoded His_6 tags to facilitate purification. The original 14B7 scFv with only a c-terminal His_6 tag was used as a hexa-histidine tagged ligand. All ligand proteins were expressed from pAK400 in *E. coli* BL21 in 250 ml cultures of TB media, induced with 1 mM IPTG for 3-5 hours at 25°C before cell harvesting and periplasmic fractionation via osmotic shock as previously described (Maynard *et al.*, 2005). Recombinant antibody-based ligands were purified via sequential immobilized Ni^{2+} affinity chromatography and SEC, using HEPES-buffered saline (HBS; 10 mM HEPES, 150 mM NaCl, pH 7.4). The MBP-based ligands were purified using an amylose affinity column and eluted with maltose containing buffer (200 mM HEPES, 150 mM NaCl, 10 mM maltose, pH 7.4), prior to SEC.

3.3.2 Library generation by CDR and random mutagenesis

The 3D5 gene was generated by total gene synthesis from published amino acid sequence (Genscript), including previously identified solubilizing mutations (referred to here as 3D5) [99] and cloned via *Sfi*I-*Sfi*I ligation into the pMoPac24 phage display vector [141], which introduces a c-myc tag at the *c*-terminus of the displayed protein. To identify clones with desired peptide specificity, two CDR libraries were generated by modified Kunkel mutagenesis under error-prone conditions: HCDR3 only and HCDR2+HCDR3 simultaneously. Randomized codons were selected to retain anti-peptide binding capabilities as described previously [159]. For generation of EE-specific

antibodies, amino acid sequencing by Edman degradation (ICMB Protein Microanalysis Facility, University of Texas at Austin) of proteolytic fragments of the commercially available GluGlu antibody (Covance), identified a candidate HCDR1 sequence. The HCDR2 loop included 13 amino acid modifications (theoretical size, 9×10^{14}), while HCDR3 loop length was set at seven residues to reflect observed diversity in anti-peptide antibodies (theoretical diversity, 6.4×10^{8}). The 3' H2 oligonucleotide sequence is CACGGTGAGTGTGGCCCTSNNCTTVNHMNYSBNGTTATASNNSNNSNNSNNA YYSNNARDMYDAATSNNTCCGATCCACTCCAGACC, while the 3' H3 oligonucleotide sequence is CCTTGACCCCAGTAATCCATAGCSNNSNNSNNGCTS NNSNNSNNSNNABNTGCACAGTAGTATACG.

A third library was generated to introduce random mutations into a pool of scFvs selected for EE-specificity. Here, 3D5 variants with desired epitope specificity were subjected to error prone PCR with Mutazyme II DNA Polymerase (Stratagene) using flanking primers (5' scback and 3' scforlong, IDT [158]) and exponential amplification according to manufacturer's instructions. Briefly, reaction mixtures were heated at 95°C for 4 min, followed by 25 cycles of incubation at 95°C for 30 sec, 52°C for 30 sec and 72°C for 1 min to introduce a predicted 3-4 mutations per 1000 bp. Gel purified PCR products (Qiagen) were used in a modified Kunkel mutagenesis step to produce two libraries. Library mutation rate and diversity was assessed by plasmid DNA sequencing at the University of Texas Core Facility using primer 5' pAKpel [158] (IDT).

3.3.3 Selection and screening by phage display

M13 phage monovalently displaying scFv-gpIII fusions were prepared as previously described [158]. After precipitation with ¼-volume PEG-2.5 M NaCl and resuspension in PBS, the phage concentration was quantified by absorbance:

$$\text{virions/ ml} = [(A_{269}-A_{320})*6*10^{16}/(\#\text{bases per virion})]. \quad \text{(Eq. 3.1)}$$

For panning, 10^{12} plaque forming units (pfu) were added to blocked ELISA wells (Costar) coated with either anti-c-myc antibody (9E10, Sigma), His6 or EE presenting ligand. After equilibration, and washing with PBS-0.05% Tween, bound phage were eluted with 0.1 M glycine-HCl pH 2.2, transferred to a new tube and neutralized with 2 M Tris, pH 7.0. Phage were then amplified in *E. coli* in preparation for the next panning round or used to infect *E. coli* and plated to isolate single clones. Panning involved two cycles, each consisting of three selection rounds: one with immobilized anti-c-myc antibody to enrich for full-length scFv and remove variants with primer encoded stop codons or frameshifts, followed by two rounds with peptide-tagged ligands.

Individual phage clones were analyzed by phage ELISA to confirm enrichment of peptide-specific clones and to screen candidates for biophysical characterization. Phage from single clones were produced in 200 µl in sterile 96-well plates (Costar). The plates were centrifuged and supernatant transferred to coated and blocked ELISA wells (Costar). After washing with PBS-0.05% Tween, bound phage were detected by anti-M13-HRP (1:2500, GE Healthcare) with tetramethylbenzidine (TMB, Sigma) substrate and the resulting absorbance at 450 nm recorded. Binding of each clone to immobilized anti-c-myc antibody (1 µg/mL), peptide ligand (MBP-EE or 14B7-His6 at 4 µg/mL) and blocked wells (5% milk) was monitored. Clones with a high ratio of peptide ligand to

anti-c-myc signal, indicating high peptide-binding specific activity, were further characterized. To rank the relative affinities of these high-activity variants, phage were produced from 100 ml cultures and the concentration of phage particles quantified by absorbance prior to ELISA analysis. The phage concentration resulting in 50% of the maximum ELISA signal (EC_{50}) were compared to select final candidates for expression and characterization as soluble scFv protein.

To confirm conversion of specificity, phage displaying the 3D5/EE_48 or commercial anti-His-HRP (Invitrogen) were used to probe a Western blot containing host proteins presenting the EE (MBP-EE, scFv-EE), His6 (scFv-His_6) or both peptides (scFv-EE_2). Phage displaying 3D5 were not used as divalent display is required, due to low affinity His6 recognition [99]. A 15% SDS-PAGE gel was loaded with 10 ul cellular lysate from cells expressing ligand, and electrophoresed prior to transfer to PVDF membrane. After blocking with 5% non-fat milk in PBS, the blot was incubated with 8 x 10^{10} virions/ml fresh 3D5/EE_48 scFv displaying phage for 1 hour at room temperature, washed three times with HBS-0.05% Tween and incubated with anti-M13-HRP secondary antibody (1:5000). Signal was developed with SuperSignal West Extended Duration substrate (Pierce), and the resulting image captured on Kodak film. The blot was stripped with mild stripping buffer (200 mM glycine pH2.2 with 0.1% w/v SDS and 1% Tween-20), re-blocked and re-probed with commercial anti-His-HRP (1:5000, Invitrogen) to detect His6-containing ligands. The blot was stripped a second time, blocked and probed with anti-MBP-HRP (1:2500, Invitrogen) to confirm similar loading between wells.

3.3.4 Chaperone protein expression, purification, and complexation

The parent 3D5 and scFv genes selected from phage display experiments were subcloned into the *Sfi*I-*Sfi*I site of pAK400 for scFv expression [158], or pMoPac54, to produce an scAb (an scFv appended with a human kappa constant domain as a convenient detection handle [141]). Protein was secreted into the bacterial periplasm of *E. coli* strain BL21, isolated by osmotic shock and purified by immobilized Ni^{2+} affinity chromatography and SEC using a Superdex S75 column (GE Healthcare), as previously described [160]. The Superdex S75 column was calibrated using a Low Molecular Mass gel filtration calibration kit (GE Healthcare).

Protein purity and size were characterized by SDS-PAGE under reducing or non-reducing conditions [161]. Protein solubility was determined by concentrating the protein to ~20 mg/mL, incubating for four days at 4°C, centrifuging for 10 min at high speed to pellet insoluble particles and quantifying the concentration of protein remaining soluble. Stability was assessed as the mid-point for thermal unfolding, using a fluorescence assay [162]. Purified protein (20 µl at 280 µg/mL) or buffer blank and Sypro Orange (1 µl of a 1:1000 dilution; Molecular Probes) were heated in a Real Time PCR instrument (7900HT Fast Real-Time PCR System (Applied Biosystems) from 20 – 85°C in increments of 0.5 °C and analyzed with SDS.2 (Applied Biosciences). The scFv monomer-to-dimer ratio was determined from SEC traces by calculating the area under the curve for each peak with Unicorn software (GE Healthcare). Protein concentration was assessed by micro-BCA assay with a BSA standard curve and buffer blank (Pierce). To facilitate direct comparisons, all 3D5 and variant characterization values reported here were performed with these methods and specific values may differ slightly from those previously reported [99].

Complex formation between 3D5/EE_48 and two ligands, scFv-EE3 and MBP-EE (see above), was assessed by SEC. Equimolar volumes of purified 3D5/EE_48 and either scFv-EE3 or MB-PEE (~1 μM each) were combined and allowed to incubate on ice for six min followed by separation using an analytical Superdex S75 column (GE Healthcare) equilibrated with 10 mM HEPES, 150 mM NaCl, pH 7.4. Fractions of interest were concentrated and characterized by SDS-PAGE. Control experiments applied the same quantity of each species alone as used in complexation experiments. The Superdex 75 column was calibrated using a Low Molecular Mass gel filtration calibration kit (GE Healthcare) supplemented with cross-linked albumin (Sigma).

3.3.5 Determination of chaperone-peptide binding affinity

Direct ELISA with purified scFv protein was performed in two orientations: scFv as an immobilized capture molecule or scAb detection of immobilized EE-tagged protein. For the former, wells were coated overnight at 4°C with 50 μg/mL scFv variant in PBS, prior to blocking with 5% milk in PBS. MBP-EE was serially diluted (1:2) from an initial concentration of 100 μg/mL, followed by washing and detection with anti-MBP-HRP (1:2500). For the inverse configuration, plates were coated with EE-tagged proteins (4 μg/mL) followed by 1:2 serial dilutions of scAb protein from 200 μg/mL. In this case, detection was achieved with anti-human-kappa-HRP (1:2500; Sigma) and TMB substrate. To assess the pH sensitivity of the binding interaction, ELISAs were performed in which the scFv-ligand interaction proceeds at pHs ranging from 6.0 to 8.0, in 0.5 increments. To rank the relative affinities, the EC_{50} concentrations were compared.

Kinetic binding assays were performed with proteins bearing *c*-terminal His_6 or EE-tags and internal EE-tags to quantify scFv peptide selectivity using a BIAcore 3000

(GE Healthcare). Peptide-binding scFv or protein ligands were coupled to CM5 chips using NHS-EDC chemistry to a level of ~500 RU. The signal from a flow cell coupled with a control scFv (14B7-His_6 [39]) was used to correct for non-specific binding to the matrix, while control scFv injections corrected for changes in sample refractive index. Soluble protein ligands were injected in a duplicate dilution series from 8 to 0.1875 μM at a flow rate of 50 μl/min to minimize mass transport effects. The association rate constant (k_{on}), dissociation rate constant (k_{off}), and equilibrium dissociation constant (K_d; $K_d = k_{off}/k_{on}$) were calculated assuming a Langmuir 1:1 binding model with BIAevaluation software. Only data sets with $\chi^2 < 0.5$ were used.

3.3.6 Protein crystallization

3D5/EE_48 was crystallized by the sitting drop vapor diffusion method at 4 °C. Conditions were optimized based on those reported for 3D5 [99]. 1-2 μL of protein solution in HBS buffer at 3.8 mg/mL chilled on ice was mixed with 1 μL sample of reservoir solution containing 0.1 M Mes (pH 6.4), 0.1 M magnesium acetate, and 20-24% (w/v) PEG 8000. Crystals of rectangular or triangular shape appeared in four days and grew to a maximal size of 40-60 μm within 4 weeks. The largest crystals grew when the reservoir to protein ratio was 1:1.33-1.66.

3.3.7 Data collection, structure determination and refinement

Crystals were harvested at 4°C and cryocooled using a solution consisting of 85.5% (v/v) reservoir solution and 14.5% (v/v) ethylene glycol. Crystallographic data were collected using a wavelength of 1 Å at the GM/CACAT beamline (Darien, IL) equipped with a 5 μm mini-beam setup. Data were processed with XDS [163] and Scala

[164]. The structure of 3D5/EE_48 was solved by molecular replacement with Molrep [164] using a polyalanine search model derived from parent 3D5 asymmetric unit (PDB ID 1KTR) from which all non-protein atoms and loop residues were removed. All four 3D5/EE_48 scFv monomers present in the asymmetric unit were identified from Molrep. The atomic model was fit to the respective electron density map using Coot [165], and then iteratively refined with Refmac [164]. After several initial rounds of refinement using tight non-crystallographic symmetry restraints, refinement including Translation/Libration/Screw motions and medium noncrystallographic restraints was conducted. Of the 947 residues present in the asymmetric unit, 99.2 % are in most favored and additional allowed regions of the Ramachandran plot. The final model has been deposited in the PDB (PDB ID 3NN8). Figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 0.99rc6, Schrödinger, LLC). Electrostatic surface potentials were calculated using APBS [166] and visualized using Python Molecular Viewer 1.5.4 [167]. Computational peptide docking was conducted with ClusPro [168].

3.4 RESULTS

3.4.1 Selection of 3D5 variants

We identified an scFv scaffold to use as a starting point for engineering peptide-binding chaperones by examining a family of structurally characterized antibody fragments binding small molecules (PDB IDs including 1KTR, 1MAJ, 2CJU, 1DLF, 2UUD, 1DSF, 1WZ1, 1N4X, 2G60) that share a highly conserved variable light chain (VL from the murine Vκ1 germline, >90% identity) and, if crystallized, a major crystal contact. One member of this family, the His6-specific 3D5 scFv, had previously been

displayed on M13 bacteriophage [100]. We hypothesized that we could enhance and/or convert scFv peptide specificity while retaining the favorable crystallization characteristics of 3D5 through CDR and random mutagenesis, coupled with a phage display selection strategy in which peptide binding affinity, solubility, stability and expression level are used as proxy variables for crystallizability. Similar scaffolding approaches have been effective for antibody humanization and thermodynamic stabilization [128, 169].

To increase the versatility of our crystallization chaperones, we sought to identify variants with affinity for either the His_6 or EE (sequence: EYMPME) hexa-peptides. The chemical diversity of the EE peptide would be expected to enhance binding interactions while the inclusion of a proline would limit conformational entropy [170]. In order to engineer scFvs with the desired peptide specificity (His_6 or EE), three libraries with randomized CDRs were generated by methods previously described [159]. Since the heavy chain (V_H) typically dominates ligand interactions [171], V_H CDRs 2 and 3 (HCDR2 and HCDR3) were randomized to convert peptide specificity while retaining the desirable crystallization properties of 3D5. The three libraries of scFv HCDR variants (actual library size ~10^7 each) were monovalently displayed on the surface of M13 phage via fusion to coat protein gpIII, and scFv variants were selected for ligand binding specificity using a series of panning cycles. First, full-length scFvs, which present a c-terminal c-myc epitope, were enriched from prematurely truncated variants resulting from oligonucleotideencoded stop codons via immunoprecipitation. Next, eluted phage were amplified in *E. coli,* and variants with desired peptide specificity selected via phage binding to an immobilized host protein presenting either the His_6 or EE peptide. One host protein, the 14B7 scFv with a terminal His_6 peptide (scFv-His_6), was employed for selection of hexa-histidine specific variants. Two ligand proteins were used to select for

EE peptide binders: maltose binding protein (MBP) with a single *c*-terminal EE tag (MBP-EE), and another scFv with two internal tandem EE tags to allow for steric accessibility within the Gly-Ser linkage between V_H and V_L immunoglobulin domains (scFv-EE_2).

The amplification and selection procedure was repeated twice, using different immobilized host proteins during each cycle to ensure selection for peptide, as opposed to host protein, specificity. Next, weakly peptide-reactive phage were pooled and subjected to random mutagenesis to yield the libraries, one based on EE-specific scFv. Sequencing of 20 individual clones from each library comprising ~10^7 unique members confirmed library diversity and the anticipated ~0.5% mutagenic rate. An additional three rounds of phage selection yielded the pool of EE peptide-specific scFv variants.

After screening several hundred clones by monoclonal phage ELISA followed by phage titration ELISA to rank clones by binding affinity, two His_6 (denoted 3D5/His_#) and six EE-specific (denoted 3D5/EE_#) scFv variants with unique sequences were identified (Table 3.1). Of these, two clones, 3D5/His_683 and 3D5/EE_48, provided the highest specific binding activity (measured as the ratio of peptide tag/anti-c-myc ELISA signal). Western blot analysis provided a clear verification of peptide specificity: 3D5/EE_48 displayed on phage bound host proteins with either internal or c-terminal EE peptides, but not those with only c-terminal His_6 peptides (Fig. 3.1a). These scFv variants were then expressed and purified as soluble protein (Fig. 3.1b and 3.1c), and characterized for binding activity by ELISA and surface plasmon resonance (SPR; Fig. 3.2a-f) analysis and for enhanced biophysical properties (see below).

Table 3.1 Comparison of 3D5 scFv CDR H3 regions

				HCDR3																
Kabat #	90	91	92	93	94	95	96	97	98	99	100	100A	100B	100C	100D	101	102	103	104	105
3D5	Y	Y	C	E	S	Q	S	G								A	Y	W	G	Q
3D5/His_683	Y	Y	C	A	A	S	S	P	Y	S	M	R	A	A	M	D	Y	W	G	Q
3D5/His_67	Y	Y	C	E	R	A	R			S	P	R	A	A	M	D	Y	W	G	Q
3D5/EE_48	Y	Y	C	A	R	R	G	G	S	S	H	Y	Y	A	M	D	Y	W	G	Q

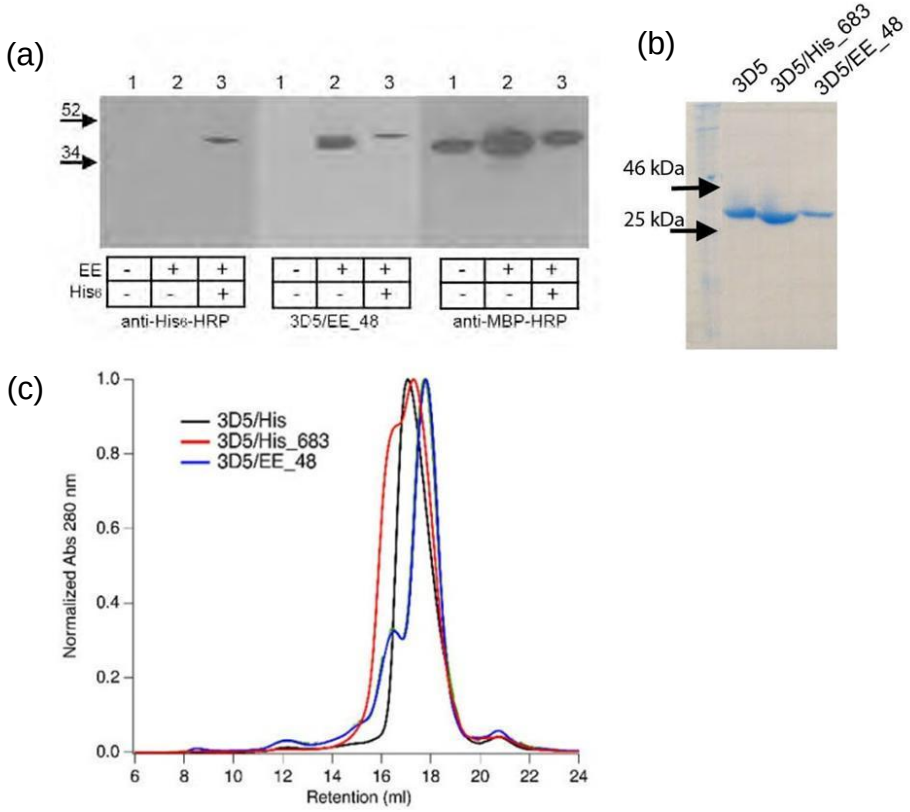


Figure 3.1. Analysis of scFvs described in this study.

(a) Western blot detection of MBP ligands by anti-His6-HRP, 3D5/EE_48 displayed on M13 phage and anti-MBP-HRP on the same blot. Lanes are 1: MBP, 2: MBP-EE, 3: MBP-EE-His_6.
(b) Size and purity by SDS-PAGE.
(c) Purification by SEC. Molecular weight standards on S200 elute at 8.6 ml (void volume; blue dextran 2000); 15.6 ml (75 kDa conalbumin); 16.5 ml (43 kDa ovalbumin); 18.6 ml (29 kDa carbonic anhydrase); 19.0 ml (13.7 kDa ribonuclease A); 21.0 ml (residual peak from Ni^{2+} affinity column).

3.4.2 Characterization of 3D5/His variants

The selected variants 3D5/His_67 and 3D5/His_683 differ from each other and 3D5 in the HCDR2 (3D5/His_67) and HCDR3 (3D5/His_67 and 683). These variants harbor longer CDR3 lengths with several amino acid differences (Table 3.1). In our hands, 3D5/His_683 expressed nearly three-fold better in *E. coli* than 3D5 (8.5 vs 3.1 mg/L culture; Table 3.2), exhibits enhanced scFv solubility (estimated as 16.6 versus 2.3 mg/ml, respectively) and modestly improved affinity (K_d 808 versus 4700 nM). At concentrations relevant to crystallization (~4 mg/ml) 3D5/His_683 elutes from a gel filtration column as a mixture of a monomer and dimer (Fig. 3.1c). In contrast, 3D5/His_67 expressed at lower levels, but exhibited similar affinity for His_6 (K_d 760nM). One of two key residues in HCDR3 that stabilize the bound His_6 in the 3D5 crystal structure, Glu93 or Ser96, is retained in each variant (Table 3.1), yet these variants and others we tested all exhibited micromolar affinity for His_6 (K_d 3-4 μM). Thus, even though these variants possess rather different HCDR3s than 3D5 and more favorable biochemical properties, their affinity for His_6 is not substantially improved over that of 3D5. These results suggest that the His_6 binding site is well-organized for peptide binding, or that HCDR3 may contribute fewer productive interactions than expected. For these reasons, plus concerns regarding the utility of His_6 as a peptide ligand, these clones were not pursued further.

Table 3.2. Biophysical characteristics of 3D5 scFv variants

	3D5	*3D5/His_683*	*3D5/EE_48*
Expression level (mg/L culture)	3.1	8.5	2.1
Solubility (mg/mL)	2.3	16.6	12.8
Melting temperature (°C)	46.5 ± 0.5	53.6 ± 0.0	47.2 ± 0.3
% AA identity	100	96.2	85
% Monomeric protein[a]	50	62	81
K_d (nM), 14B7-His_6 (χ^2)	4700 (0.08)	3200 (0.26)	ND[b]
K_d (nM), MBP-EE (χ^2)	ND[b]	ND[b]	767 (0.03)

[a]% Monomer measured upon initial purification
[b]ND = not detected

3.4.3 Characterization of 3D5/EE_48

The lead 3D5/EE scFv candidate, 3D5/EE_48, retains 85% amino acid identity relative to 3D5. In addition to novel HCDR sequences (Table 3.1), two key amino acid changes in the V_H framework identified during random mutagenesis (E6Q and S74T) were instrumental in improving scFv expression and affinity. The impact of Glu versus Gln at position six has been previously described [172].

The 3D5/EE_48 scFv displayed no detectable binding affinity for *c*-terminal His_6 tags and instead is able to bind both *c*-terminal and internal EE-tags with similar affinities, K_d 389 and 212nM, respectively (Figures 3.1a, 3.2c, d, e; Tables 3.2, 3.3). A terminal EE tag followed by a His tag was recognized with higher affinity than a naked EE tag (K_d 389 versus 767 nM), perhaps to due protease protection and reduced entropy with an additional c-terminal extension. Varying the pH in 0.5 increments from 6.0 to 8.0 or increasing the number of internal EE tag repeats from two to three had no detectable effect on affinity as measured by ELISA (data not shown) and SPR (K_d 25-30 nM; Table 3). An increase in affinity was observed for ligands harboring multiple E repeats versus single repeats, likely due to re-binding effects, as the measured on-rates are similar but

the off-rates are slower (Table 3.3). Use of 3D5/EE_48 displaying phage as detection reagents in ELISA (data not shown) and Western blot (Fig. 3.1a) demonstrated specific binding of EE but not His_6-tagged ligands. Expression levels of 3D5/EE_48 (2.1 mg/L culture) are similar to that observed for 3D5 (3.1 mg/L culture), but the solubility increased from 2.3 mg/ml to >12 mg/ml. In addition, 3D5/EE_48 is initially purified as a predominantly monomeric species (~80% of total eluted protein; Table 3.2) and retains this monomeric state when concentrated up to at least 3 mg/ml (Fig. 3.1c). This contrasts with the lower initial ratio of monomeric to dimeric protein (Table 3.2) and slow conversion of purified monomer to dimer observed for 3D5 and 3D5/His_683 under similar conditions (Fig. 3.1c). The melting temperature of 3D5/EE_48, is almost identical to 3D5, indicating similar thermal stabilities (Table 3.2). Overall, 3D5/EE_48 exhibits similar or enhanced biophysical properties as compared to 3D5 in terms of affinity, expression level, solubility, stability, and homogeneity of oligomerization state.

Table 3.3 Characterization of 3D5/EE_48 scFv binding kinetics by SPR

ligand	k_{on} (1/Ms)	SD (%)	k_{off} (1/s)	SD (%)	K_d (nM)	χ^2
MBP-EE	1.5E+05	62	7.3 E-02	24	767	0.030
scFv-EE_2	3.9E+05	30	9.0 E-03	34	25.5	0.586
scFv-EE_3	3.6E+05	37	9.2 E-03	41	29.6	0.598

3.4.4 Complexation of 3D5/EE_48 with EE-tagged proteins

The ability to isolate complexes of 3D5/EE_48 and client proteins expressing the EE tag was assessed next using size exclusion chromatography (SEC). Equimolar concentrations (~1 μM) of 3D5/EE_48 and the client protein were combined and fractionated using an analytical gel filtration column. Fractions corresponding to the

eluted peaks were analyzed by SDS-PAGE and compared to control runs with isolated binding partners. One client protein, scFv-EE_3 used originally for selections (see above), elutes as a dimer, and runs slightly higher than its expected molecular mass by SDSPAGE, likely due to an extended conformation of the individual V_H and V_L domains within the monomer (Fig. 3.3a). Complexation with 3D5/EE_48 results in a single elution peak that corresponds to a molecular mass consistent with a heterotetramer, i.e., an scFv-EE_3 dimer with two bound 3D5/EE_48 monomers (Fig. 3.3a). The second client protein tested was MBP-EE, which harbors only a *c*-terminal EE tag. Although MBP-EE by itself elutes as a monomer, complexation with 3D5/EE_48 yields two higher molecular weight complexes with molecular masses consistent with a heterodimer and heterotetramer (Fig. 3.3b). Given the lack of dimerization precedent for MBP, the heterotetramer could arise from a domain swapped 3D5/EE_48 dimer in which two distinct binding sites for MBP-EE are presented. Domain swapping has been proposed as a mode for 3D5 dimerization [99].

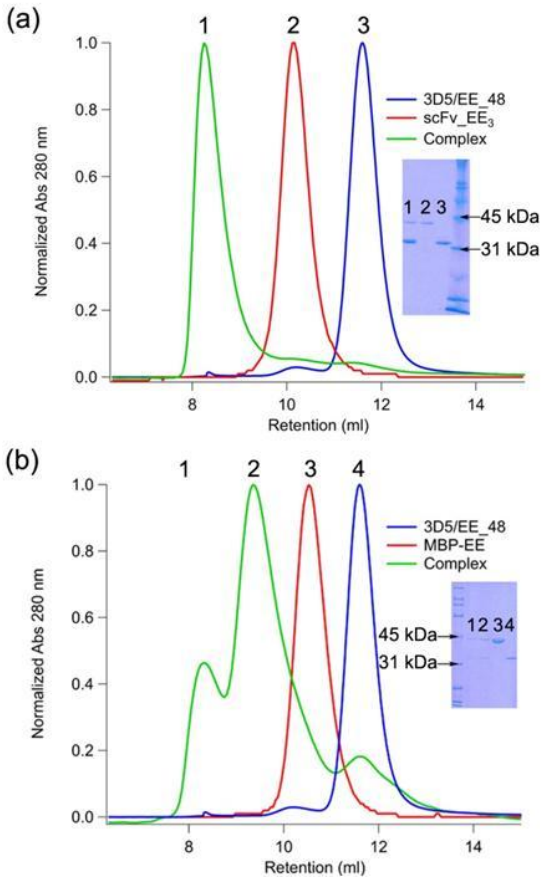


Figure 3.2 Isolation of 3D5/EE_48 complexes with EE-tagged client proteins by Size Exclusion Chromatography (SEC).

(a) 3D5/EE_48 incubated with scFv-EE3. Elution peak 1 corresponds closely to the expected retention volume of a heterodimer complex as described in text.
(b) 3D5/EE_48 incubated with MBP-EE. Elution peaks 1, 2 correspond to heterotetramer and heterodimer complexes as described in text.
For (a) and (b), SDS-PAGE analysis of elution peaks confirms protein and complex identities. In addition, individual elution profiles are overlayed. Red: scFv-EE3 or MBP-EE. Blue: 3D5/EE_48. Molecular weight standards elute on S75 at 7.4 ml (void volume; dextran 2000); 8.9 ml (132 kDa albumin dimer); 9.6 ml (75 kDa conalbumin); 10.9 ml (44 kDa ovalbumin); 12.7 ml (29 kDa carbonic anhydrase).

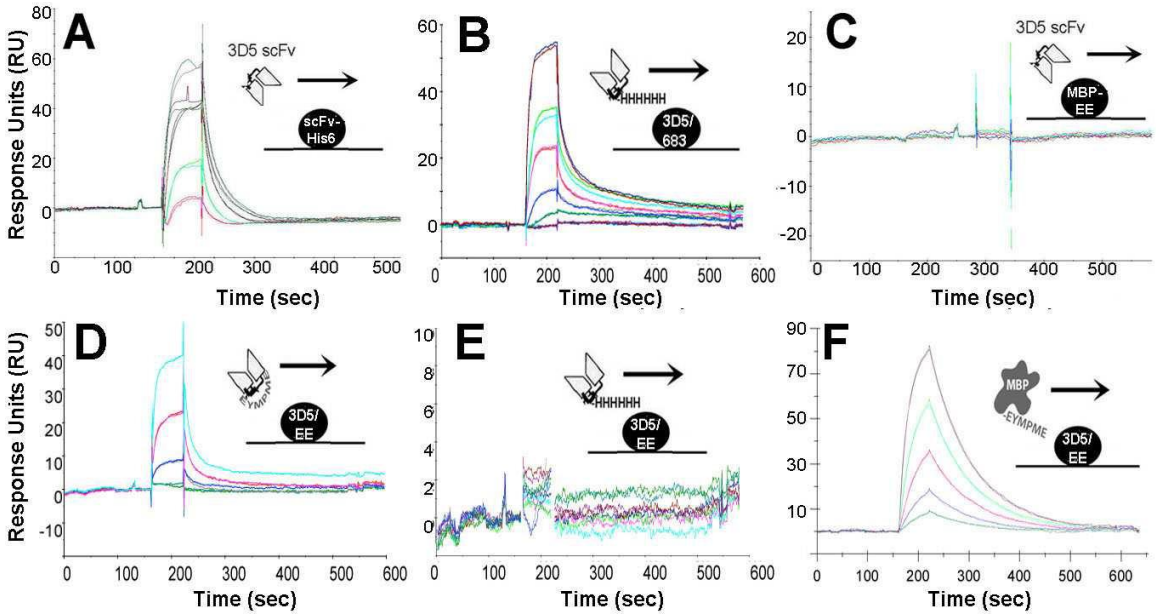


Figure. 3.3 Peptide-scFv binding kinetics.

Top row, analysis of 3D5 and 3D5/His_683 binding affinity and specificity His_6 by SPR. Binding partners were injected in duplicate in concentrations ranging from 8000 to 125 nM.

(a) 3D5 scFv recognition of immobilized scFv-His_6 ligand;

(b) binding of scFv-His_6 ligand to immobilized 3D5/His_683;

(c), 3D5 recognition of control MBP-EE ligand. Bottom row, analysis of 3D5/EE_48 affinity and specificity for EE-tagged proteins by SPR. Binding of

(d), a single, internal EE peptide presented by scFv-EE1,

(e), a control scFv lacking an EE tag, scFv-His_6 (buffer spikes erased for clarity); and

(f), a c-terminal tag presented by MBP-EE-His6 to immobilized 3D5/EE_48 scFv ligands. Ligands were injected at concentrations ranging from 6000-475 nM for MBP-EE, 2000-125 nM for the scFv ligands. Cartoons illustrate the orientations and identity of binding partners.

3.4.5 Structure of 3D5/EE_48

Crystals of 3D5/EE_48, grown under conditions used to crystallize 3D5 [99], appeared within four days and continued to grow over several weeks. The structure of 3D5/EE_48 was solved by molecular replacement using a search model derived from the 3D5 coordinates (see Methods, Table 3.4). Although the crystals were grown under

similar conditions, and the proteins share a high level of sequence identity, the two scFvs do not crystallize in the same manner (Fig. 3.4). First, whereas 3D5 crystals belong to a trigonal space group ($P3_221$), crystals of 3D5/EE_48 belong to a cubic space group (F23). The asymmetric unit of 3D5/EE_48 contains four molecules whereas 3D5 contains one V_H-V_L pair (Fig. 3.4a). In addition, in spite of the fact that no amino acid changes occurred in the major 3D5 crystal contact, the contact is not preserved in the 3D5/EE_48 lattice. Whereas the crystal lattice of 3D5 is built by alternating V_H/V_L subunits from neighboring molecules, that of 3D5/EE_48 relies primarily on HCDR residues from adjacent molecules (see arrows, Fig 3.4b). The second largest contact in the 3D5 crystal lattice (305 $Å^2$) has become the largest crystal contact (560 $Å^2$) in the 3D5/EE_48 lattice with several additional hydrogen bonds formed at this interface (see shaded area, Fig. 3.4b). Finally, as a consequence of lattice changes, 3D5/EE_48 crystals consist of 66% solvent with a channel ~52 Å wide while 3D5 crystals consist of 77% solvent and a channel ~70 Å wide (Fig. 3.4c).

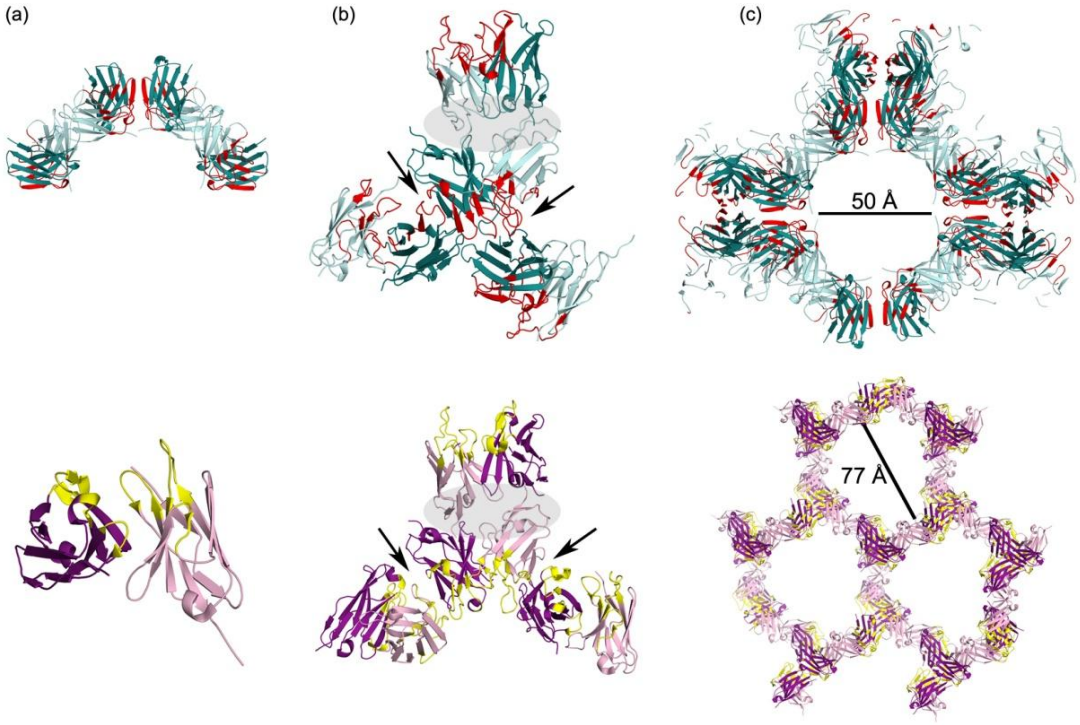


Figure 3.4 Comparison of 3D5/EE_48 (top) and 3D5 (bottom) crystal lattices.

(a) Asymmetric units.
(b) Crystal contacts. The preserved contact common to both lattice networks shaded grey. New crystal contact comparison depicted in arrows and labeled.
(c) Lattice structure with size of central cavity indicated. For both structures, lighter hue indicates the VL chain while the darker hue indicates VH. CDRs for 3D5/EE_48 are depicted in red and those for 3D5 are yellow.

The overall structure of the 3D5/EE_48 scFv remains very similar to that of the parent 3D5 (average rmsd ~0.55 Å for main chain atoms in V_L and ~1 Å for VH domains); however, changes observed in the CDR regions reconfigure the peptide binding region to accommodate an EE-tag (Fig. 3.5). In the V_L CDR1 loop (LCDR1) of 3D5/EE_48, slight movement in residues His27d-Asn30 may be influenced by the presence of neighboring Leu93 in the V_L CDR3 loop (LCDR3), instead of the corresponding His residue at this position in 3D5 (Fig. 3.5a). Another substitution in

LCDR3 of 3D5/EE_48, introduction of a Pro96 for the Phe in this position in 3D5, appears to open up the peptide binding groove to accommodate longer peptides, and in particular, may allow internal peptides to be recognized (Fig. 3.5a). Compared to the LCDRs, HCDRs are more divergent both in sequence and in structure. In 3D5/EE_48, the beta-hairpin in HCDR2 as a whole shifts closer to HCDR1. HCDR3 differs primarily in its longer length, which significantly alters the shape of the peptide binding region when compared to 3D5. The binding surface near the interface of the heavy and light chains forms a pronounced tri-lobed hydrophobic pocket (Fig. 3.5b and 3.5c). The electrostatic surface potential reflects a charge distribution complementary to that of the peptide in this region (Fig 3.5b).

Table 3.4 Data collection and refinement statistics

	3D5/EE_48
Data collection	
Space group	F23
Cell dimensions	
a=*b*=*c* (Å)	266.64
α=β= γ (°)	90
Resolution (Å)*	154.3-3.1 (3.18-3.10)
R_{sym}	12.4 (34.2)
I/ σI	5.6 (2.1)
Completeness (%)+	99.6 (99.6)
Redundancy	5.6
Refinement	
Resolution (Å)	154.3-3.1
No. reflections	26932
R_{work} / R_{free} [a]	18.3/23.7
No. molecules	
Protein residues	947
B-factor protein (A^2)	40.8
R.m.s deviations	
Bond lengths (Å)	0.016
Bond angles (°)	1.871

[a] Highest resolution shell is shown in parenthesis. 5% of reflections were selected for R_{free}.

We turned to computational docking to model EE-peptide binding to 3D5/EE_48 (Fig. 3.5c) because no crystals of adequate size for structure determination containing both 3D5/EE_48 and EE-peptide have been grown to date, and soaking with the commercial EE-peptide (Covance) has not yielded crystals with bound peptide. The EE peptide is predicted to bind in an orientation in which the central proline (Pro4) introduces a kink, allowing peptide residues Tyr2 and Met3, to reach into the hydrophobic binding pocket. In this working model, V_H residue His50 appears to stabilize peptide Tyr2 through hydrophobic interactions, while V_H residue Arg95 forms key polar interactions with multiple peptide side chains (Tyr2, Met3, Glu6). Peptide residues Glu1 and Glu6 stabilize this binding mode through surface electrostatic interactions, and hydrogen bonding interactions between the peptide backbone amide and carbonyl stabilize the peptide conformation. In the case of a terminal EE tag, the C-terminus may compete for the Glu6 side chain interactions, and/or more flexibility of the tag may destabilize the peptide backbone interactions. In this docked model, peptide residue Met5 does not appear to be directly recognized by 3D5/EE_48. Notably, V_L residues predicted to form interactions with peptide are conserved between 3D5 and 3D5/EE_48, while V_H residues contributing to peptide interactions, such as Arg95, were altered during engineering, underscoring the role of the V_H in peptide recognition.

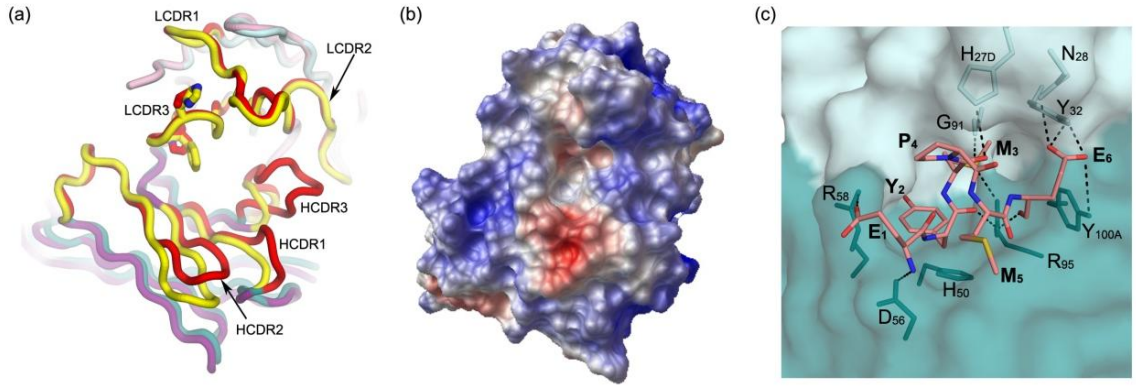


Figure 3.5 Analysis of 3D5/EE_48 structure.

(a) Superimposition of 3D5/EE_48 and 3D5 with CDRs labeled. Amino acid changes discussed in the text are represented as ball-and-stick. LCDR1, LCDR2, LCDR3 indicate VL CDR loops 1, 2, and 3 respectively. HCDR1, HCDR2, HCDR3 indicate VH CDR loops 1, 2, 3, respectively.
(b) Calculated electrostatic surface potential of the binding site (-15 kT to +15 kT).
(c) Modeled peptide binding interaction of EE peptide with 3D5/EE_48. Color scheme identical to Fig. 3.

3.5 DISCUSSION

Crystallization chaperones are proposed to aid co-crystallization by several distinct mechanisms, including immobilizing flexible regions, concealing exposed hydrophobic regions, and providing polar surfaces capable of forming lattice contacts [79, 86]. To date, most co-crystal structures have employed antibody fragments because the molecular requirements for ligand binding are well understood and their hypervariable regions can be modified to recognize nearly any epitope of interest [159, 173].

Typically, antibodies that recognize specific epitopes on unmodified target proteins are identified through traditional hybridoma screening or library selection

techniques [91, 98, 174, 175] where there is minimal control over the epitope recognized. Moreover, the identification and optimization of a chaperone tailored to each client protein of interest is an expensive and time-consuming process. An attractive alternative is the use of commercially available purified monoclonal antibodies for common epitopes, such as commercially available anti-His antibodies. Unfortunately, the hybridomas secreting these antibodies are not available and the cost to purchase purified antibody sufficient for use in crystallization trials is prohibitive. Moreover, without the gene sequence available, the biophysical properties and format of the antibody (*e.g.*, scFv, scAb, Fab) cannot be readily altered. Finally, even in the case where sequences are known and genes for the corresponding antibody fragments can be synthesized for recombinant expression, antibody fragments often express with relatively low yield in *E. coli* and lack suitable solubility and stability profiles.

The engineered scFv chaperone approach complements non-antibody formats that have been developed to allow modular recognition of a specific binding partner (*e.g.,* DARPin, V_{HH}) [80, 85, 91, 92]. Whereas these alternative frameworks express in very high levels (up to 200 mg/L in the bacterial cytoplasm for DARPins) and possess a stable structure, a potential disadvantage is their small size ~15 kDa, which limits the hydrophilic surface area available for generating protein-protein crystal contacts. By contrast, scFvs are nearly twice as large, and can be readily converted to a ~50 kDa Fab format to accommodate larger client proteins with a larger hydrophobic surface area.

Our engineered scFv chaperones, derived from the previously crystallized 3D5 scFv framework and binding short His_6 or EE peptide sequences, overcome several of the aforementioned limitations of antibody fragments and represent a potentially generalizable solution to the production of high affinity protein complexes for crystallization of difficult proteins. We overcame the affinity, pH sensitivity and

solubility limitations specific to 3D5 by employing a two-step protein engineering process of randomizing the HCDR 2 and 3, followed by random mutagenesis of the selected scFvs. This selection scheme does not directly select for the ability to crystallize, as there is no clear biophysical correlate of crystallization propensity, but can select for "well-behaved" proteins, as evidenced by the increased expression levels, solubility and peptide binding affinity of our characterized variants. The initial library design focused on the HCDRs because these can be sufficient to confer high affinity and specificity [104, 176] and in the 3D5 family of antibody fragments, the VL domains are highly conserved.

After limited success in improving the biochemical characteristics of His6-specific scFvs, we converted 3D5 to EE epitope specificity. The 3D5/EE_48 scFv is expressed in high yield in *E. coli*, is highly soluble, is predominantly monomer, and is readily crystallized. The affinity of 3D5/EE_48 for internal EE-tags (K_d 212 for single, 26nM for multiple peptide insertions) likely reflects the combined effects of restricted conformational variability due to the presence of a proline in the EE-peptide, as well as the ability of the remaining peptide residues to participate in hydrogen bonding, electrostatic or hydrophobic interactions. Combined with the general reduced entropic costs of binding an internal peptide this scenario represents a desirable binding configuration for crystallization chaperone and tagged client protein. Indeed, complexes of 3D5/EE_48 with host proteins are sufficiently tight to withstand separation by SEC. In the context of a co-crystallization experiment, a modest 5 mg/ml concentration of a 30 kDa scFv chaperone protein equates to 170 μM, which is nearly 1000-fold above the measured equilibrium dissociation constant and will drive complex formation within the crystallization drop.

3.6 CONCLUSION

Unexpectedly, the crystal lattice of 3D5, whose open framework and limited use of CDRs in crystal contacts was an initial design criterion, was not preserved in 3D5/EE_48. Although the use of CDR residues in crystal contacts appears to render 3D5/EE_48 not ideal for cocrystallization, none of the residues participating in the major crystal contact of 3D5 have in fact been altered. Thus, it should be possible for 3D5/EE_48 to revert back the 3D5 lattice framework when most CDR residues are participating in a complex and CDR-based crystal contacts are no longer accessible. We are optimistic about the prospect of the ability 3D5/EE_48 to promote crystallization of "difficult" proteins, either by mediating formation of crystal contacts (as observed for KcsA [86]) or by immobilizing flexible loops (as observed for GPCRs [94, 153]). Our current efforts are focused on further engineering of 3D5/EE_48 to render the CDR crystal contact less favorable than those found in 3D5. We are also cocrystallizing 3D5/EE_48 with MBP and candidate membrane proteins in which the EE peptide has been installed into an accessible but functionally silent location. In the long term, we plan to extend our approach to generate 3D5-based scFvs or Fabs that recognize other peptide sequences, leading to a toolbox of peptide binding crystallization chaperones with homotypic crystal contacts that could be used to crystallize any protein of interest.

Accession numbers
The coordinates of the 3D5/EE_48 structure are deposited in the Protein Data Bank under PDB ID 3NN8.

Acknowledgements
We thank Benjamin Roy for the scFV-EE_1 constructs. We thank Dr. Raquel Lieberman for the SPP and SPP-EE constructs and their assistance with chaperone crystallization.

CHAPTER FOUR: Optimization of Peptide Placement in Target Proteins for chaperone assisted crystallization.

4.1 CHAPTER SUMMARY

An innovative and potentially general approach to crystallization of membrane and other hard-to-crystallize proteins employs complex formation between a protein of interest harboring a short, surface exposed peptide and hyper-crystallizable antibody fragments with peptide specificity. The antibody provides additional hydrophilic surface regions to mediate formation of crystal contacts and facilitate crystal growth. Once developed, the crystal chaperone can be used without further modification to co-crystallize multiple proteins with efforts primarily focused on peptide insertion. In order to minimize the subsequent optimization of each construct, we aim here to develop general design rules for peptide insertion into target proteins in three locations: (1) terminal, (2) inter-domain linkers and (3) internal loops. The EE peptide was introduced into three model proteins at these sites to identify constructs meeting the dual goals of rigid complex formation and high affinity antibody binding. Constructs utilizing peptide insertion at the c-terminal region are preferred for proteins with no structural information, however, a protective tail is required to reduce conformational entropy. For proteins with some structural information, single EE peptide insertion is favored at the start of the interdomain linker with four flanking residues preceding the tag resulting in higher affinity (K_D~200nM) and minimal impact to the structural stability of the target protein. Peptide insertion into linker regions connecting two domains conferred higher affinity than at protein termini (26 vs. 389 nM). When the EE peptide was introduced into a natural internal loop of maltose binding protein, however the affinity was reduced, even when flexible adapters were added to either side of the peptide (K_D = 800-1000 vs. 389

for terminal peptide). These engineered variants provide us with better guidelines for peptide insertion, transferable across multiple target proteins of interest, to facilitate rapid construction of clones to assess in complex crystallization trials.

4.2 INTRODUCTION

4.2.1 The importance of chaperone-assisted crystallization

Crystallization remains the bottleneck to obtaining structural visualization of protein using X-ray Diffraction. X-ray crystallography is the major technique used to establish the three-dimensional structure of large target protein (>30kDa). The atomic resolution image reveals protein mechanism and clarifies the role of protein structure in health and disease, resulting in organized and systematic approaches to drug design. Currently one successful technique for acquiring high resolution membrane protein structure visualization involves co-crystallization with a protein-specific chaperone for X-ray diffraction. Developments in this field have led to chaperone options for co-crystallization, which include, but are not limited to, fusion proteins like MBP and Lysozyme [61, 177], affibodies [90], VHH Camelid Domains [85], designed ankyrin repeat proteins (DARPins) [91, 92], in addition to the most popular chaperone - antibody fragments [86, 88, 89]. Fabs (fragment antigen binding) and the Fvs (fragment variable) are the most successful co-crystallization chaperones to date, having successfully crystallized many membrane proteins including, KcsA potassium channel [98, 178], Cytochrome c oxidase (COX) [79], cytochrom bc_1 complex (QCR) [152], and Na+ - citrate cymporter CitS [91] just to name a few. The recent breakthrough of the structural determination G-protein coupled receptor with parallel crystallization utilizing two

different chaperones: a targeted Fab and a T4 Lysozyme inserted into a flexible loop, demonstrates the vital role co-crystallization plays in proteonomics [61, 93].

Membrane proteins are extremely difficult to crystallize as a result of their conformational complexity. The relatively small polar surfaces available for crystal contact formation in addition to flexible intracellular loop regions and extended c-terminus interfere with homogenous lattice formation. Long linkers are particularly detrimental to crystallization [154], so it is critical that these areas are removed or restricted to produce high diffraction quality crystals. The primary role of a co-crystallizing chaperone is creating a stable complex to facilitate ordered packing. The chaperones provide improved crystallization properties for the complex by i) protecting the hydrophobic regions of the membrane protein – preventing numerous protein-protein interactions (i.e. more homogenous monomeric solution) ii) "locking down" flexible regions, minimizing conformational entropy and ii) providing additional polar surfaces for crystal contact formation (i.e better probability for ordered crystals) [7, 80, 85]. The mediated crystal contacts allow target proteins to remain suspended within the resulting lattice scaffold [79, 86]. In addition to facilitating crystallization, the chaperones provide an additional benefit by assisting with the crystallographic phase and diffraction data analysis since the three-dimensional molecular structure of these helpers is usually predetermined. Key characteristics of a successful chaperone involve the ability to bind with high affinity to the native conformation of the target protein, in addition to forming a stable and rigid complex.

4.2.2 Utility of a generic chaperone: The development of peptide-specific chaperones

With more urgency being placed on obtaining structural analysis of membrane proteins, it is critical to create a rapid, efficient and general system for co-crystallization. The current technology in co-crystallization utilizes protein-specific chaperones that bind random epitopes of the target protein [90, 179]. Recombinant in vitro antibody technology has provided tools that have improved the speed in generating binders specific for target proteins, which involves proteolysis of monoclonal antibodies [91]. This technology facilitates the selection of a chaperone, identification and optimization of a chaperone tailored to a particular protein by efficiently generating numerous individual libraries. Even with the successful selection of the antibody chaperone with favorable biophysical characteristics, the crystallization conditions of these chaperones are unknown and can result in thousands of screens that are expensive and time-consuming process [91, 174].

To bypass some of the challenges associated with current protein specific chaperones, a hyper-crystallizable peptide-specific antibody chaperone is in development. This platform allows us to easily insert our peptide into functionally silent regions of the target protein and complex the engineered antibody fragments with high affinity to short peptide epitopes and hyper-crystallization qualities for assisted crystallization. Using this process, we eliminate the need to screen large libraries using traditional hybridoma screening of library selection techniques as this peptide specific antibody becomes a universal chaperone. Control over epitope recognition is also better established as can define the peptide insertion location and are no longer restricted by the outputs of the libraries. The system also reduces the extent of crystallization condition screening required by providing a starting crystallization condition based on the non-complexed

chaperone's crystallization characteristics. A hyper-crystallizable single chain fragment variable (scFv) antibody has been successfully engineered with high affinity for the EE tag (EYMPME) [chapter three]. This antibody has shown potential as a co-crystallization chaperone with high peptide affinity and a lattice structure that contains central cavities. While high affinity has been established (26nM) it has not yet been determined if the location of the peptide is optimal for co-crystallization. Here we explore the insertion of various peptide tags at a myriad of locations, with strong focus on the optimization of the EE tag, to establish the optimal region for use in co-crystallization. Additional tags included in our studies are the Hexa-histidine (His_6) tag, commonly used in purification methods, and a substitute Flag tag (MDYKAFDNL), that had 10-fold higher affinity to the M2 antibody than the original Flag tag (DYKDDDDK) [180].

4.2.3 Importance of peptide optimization

Ligand engineering plays a critical role in our approach to a general platform for a peptide-specific co-crystallization chaperone. The limiting factor to this technology is generating a functional target construct with a presentable peptide tag. The crystallization of the GPCR β_2-Adrenergic receptor shows the laborious effort involved with properly inserting a T4 Lysozyme into a flexible intracellular loop with minimal impact on structure [93, 94]. In order to reduce the effort involved and facilitate the process, peptide placement rules need to be defined. Before optimizing location, interaction of the chaperone and the peptide must be well characterized to ensure correct conformation for peptide presentation. The sequence of the peptide not only affects the conformational structure of the tag but use of select charged residues with tendency for h-bonding and electrostatic interactions can improve the affinity. As previously discussed,

the peptide tags in this paper were selected based on its propensity for h-bonding (tyrosyine) [104], electrostatic interactions (glutamates, aspartates, lysines) [87] and restricted conformation (proline). These key residues will provide high affinity binding complexes needed for successful co-crystallization.

Peptide location whether at the terminal ends or at internal regions of the protein can impact the accessibility and rigidity of the peptide. Select peptide tags are only accessible for interaction at the c-terminal ends [99, 100]. It is therefore important to find a ligand with the ability to produce recognition and high affinity in both types of locations. The amount of rigidity contributed by the peptide will play a critical role during the formation of homogeneous complex for co-crystallization. The application of this system is widespread and can be extended to other applications including protein scaffolds for enzyme assemblies [181], however determining optimal locations in soluble regions remote from the active site of the target protein without crystal structures will be the major obstacle. Selecting these regions has always been a major obstacle in protein crystallization. Current crystallization techniques that reduce the conformational entropy by removing flexible loops without structural information have started to address these challenges. Previously this has been done with the assistance of surface topology experiments, such as Fluorescence resonance energy transfer (FRET) [182] and analysis of other proteins in the same protein class [61].

Utilizing select model proteins (maltose binding protein, scFv, TCRs and a select membrane protein-signal peptidase), extensive peptide insertions will be introduced at various locations. Observations involving engineered ligands harboring the cognate peptide will provide design rules for peptide insertion in our platform technology. Through analysis of interaction kinetics, isolated complexes and co-crystallization trials will offer guidelines for effective peptide placement that will provide high affinity and

accessibility while retaining functionality and conformational rigidity to provide homogeneous complexes for crystallization.

4.3 MATERIAL AND METHODS

4.3.1 Molecular biology of peptide-presenting proteins

Peptide sequences were incorporated into target proteins via site-directed mutagenesis with mutagenic oligonucleotides (Integrated DNA Technologies). To generate ligands with *c*-terminal peptide tags, maltose binding protein (MBP) was amplified from the *E. coli* genome, appended with either an EE (EYMPME) or 10X FLAG tag (MDYKAFDNL) and cloned into the pAK400 vector via SfiI-SfiI sites, upstream of a His_6 tag [158]. A stop codon was inserted before the vector encoded His_6 tag to generate an MBP-EE ligand without a His_6 tag. To generate ligands with internal EE tags, three model proteins were used, MBP, two-domain antibody (scFv) and two-domain T-Cell Receptor (TCR). To mimic peptide insertion into an internal loop, the EE sequence replaced a naturally occurring loop in MBP between amino acids 171 – 176 with sequence KYENGKY. Three variants were created to assess requirements for flexibility and accessibility. The entire seven amino acid loop sequence was removed and replaced with EE tag in different conformations (Table 4.1). To mimic peptide insertion into natural linker regions present in multi-domain proteins, tandem repeats of the EE coding sequence were inserted into the flexible linker connecting the two domains of the DO11.10 TCR gene (DO11-v_1,v_2), 14B7 scFv gene[39] (R30A-EE_2) or a non-native scFv consisting of the 3D5 light chain and 14B7 heavy chain (62-EE_3). DO11-e3 was engineered with the full original EE peptide (SREEEEYMPME) [106]. The number of internal EE repeats was varied from single to triplet repeats to account for steric

accessibility (Table 4.2). All engineered proteins were cloned into pak400 and, unless specified, contain a c-terminal, vector encoded His_6 tag to facilitate purification. For the MBPs with internalized EE tags, an additional c-terminal 10X FLAG tag (MDYKAFDNL) was appended prior to the vector encoded His_6 tag. Three hexa-histidine tagged ligands were employed: unmodified 14B7 scFv, DO11 TCR and MBP each with a c-terminal His_6 tag.

Table 4.1 Peptide Insertion in the Interloop of Maltose Binding Protein (171-176)

EE-tagged protein	Sequence starting at 171 in Maltose Binding Protein
MBP-v_1	GGSEYMPMEGGS
MBP-v_2	EYMPMEY
MBP-v_3	KEYMPME

Table 4.2 Peptide Insertion in the Interdomain of various antibodies

EE-tagged protein	Antibody Class	Sequence immediately following the first domain and prior to the Glycine Linker $[G_4S]_4$
DO11-e3	TCR	SREEEEYMPME
DO11-v_1	TCR	GSGSEYMPME
DO11-v_2	TCR	GSGSEYMPMEGGGGSEYMPME
R30A-EE	scFv	EYMPME
R30A-EE_2	scFv	EYMPMEGGGGTKLELKREYMPME
62-EE_3	scFv	EYMPMEGGGGTKLELKREYMPMEGGGGTEYMPME

4.3.2 Expression and purification of peptide-presenting proteins

All ligand proteins were expressed in *E. coli* strain BL21 in 250 ml cultures of TB media, induced with 1 mM IPTG for 5 hours before cell harvesting and periplasmic fractionation via osmotic shock as previously described[160]. Recombinant peptide tagged proteins were purified via immobilized Ni^{2+} affinity chromatography followed immediately by Size Exclusion Chromatography (SEC) on a Superdex 75 column in HEPES-buffered saline (HBS; 10 mM HEPES, 150 mM NaCl, pH 7.4). The MBP-based

ligand lacking a His_6 tag was purified via amylose affinity column and eluted with maltose-containing buffer (200 mM HEPES, 150 mM NaCl, 10 mM maltose, pH 7.4). The model membrane proteins, signal peptide peptidase (SPP) and a SPP variant with an internalized EE tag, with its running buffer were a gift of Dr. Raquel Lieberman (School of Chemistry and Biochemistry and Institute for Bioscience and Bioengineering, Georgia Institute of Technology).

4.3.3 Expression and purification of the chaperone protein

Chaperone 3D5/EE_48 was cloned into plasmid pAK400 for scFv expression [158], or pMoPac54, to produce a scAb format (an scFv appended with a human kappa constant domain as a convenient detection handle [141]). Protein was expressed, isolated and purified by immobilized Ni^{2+} affinity chromatography and SEC using a Superdex S75 column (GE Healthcare), as previously described [160]. The Superdex 75 column was calibrated using a Low Molecular Mass gel filtration calibration kit (GE Healthcare). Commercial M2 IgG (Sigma-Aldrich), Glu-glu Monoclonal IgG (Covance) antibodies were utilized for a control chaperone to the flag and EE peptide tags.

4.3.4 Protein biophysical characterization

A 12% SDS-PAGE (under reducing or non-reducing conditions) was used to characterize protein purity and size [161]. Protein solubility was determined by quantifying the concentration of protein in solution post-concentration to ~20 mg/mL, incubation (four days at 4°C), with centrifugation (10 min, 13Krpm) to pellet insoluble particles. Stability was assessed by quantifying the mid-point of thermal unfolding, using a fluorescence assay [162]. Purified protein (20 µl at 280 µg/mL) or buffer blank and

Sypro Orange (1 µl of a 1:1000 dilution; Molecular Probes) were heated in a Real Time PCR instrument (7900HT Fast Real-Time PCR System (Applied Biosystems) from 20 – 85°C in increments of 0.5°C and analyzed with SDS.2 (Applied Biosciences). The protein monomer-to-dimer ratio was determined by calculating the area under the SEC trace curve for each peak with Unicorn software (GE Healthcare). BCA assay with a BSA standard curve and buffer blank (Pierce) was used to assess protein concentrations.

4.3.5 Determination of chaperone-peptide binding affinity, kinetic analysis and complexation.

Direct ELISA with the chaperone scFv was performed using two different systems: purified scFv as an immobilized capture molecule or scFv presenting phage with immobilized EE-tagged protein. For the former, wells were coated overnight at 4°C with 50 µg/mL 3D5/EE_48 scFv in PBS, prior to blocking with 5% milk in PBS. MBP EE-tagged proteins were serially diluted (1:2) from an initial concentration of 200 µg/mL, followed by washing and detection with anti- MBP-HRP (1:2500). For the inverse configuration, plates were coated with EE-tagged proteins (10 µg/mL) followed by 1:3 serial dilutions of purified phage at a concentration of 10^{12} pfu/ml. Phage expression and purification was carried out following previous described methods (Chapter 3). After washing with PBS-0.05% Tween, bound phage were detected by anti-M13-HRP (1:2500, GE Healthcare) with tetramethylbenzidine (TMB, Sigma) substrate and the resulting absorbance at 450 nm recorded. Relative affinities were then ranked by comparing the concentration of EC_{50}.

Kinetic binding assays were performed with proteins displaying *c*-terminal His_6, EE or Flag-tags and internal EE-tags to quantify peptide accessibility and affinity using a BIAcore 3000 (GE Healthcare). Primarily two systems were used: first the chaperone

scFv (3D5/EE_48) was coupled to the chip with ligands flow through, the second coupled the MBP with all three tags displayed (internalized EE, sequential c-terminal Flag-His_6 tag. All proteins were coupled to the CM5 chip using NHS-EDC chemistry to a level of ~500 RU. For the first system, the signal from a flow cell coupled with a control scFv (14B7-His6 43) was used to correct for responses due to changes in sample refractive index while the second utilized a control MBP (no peptide tags). Soluble flow through proteins were injected (one minute with a 5 min dissociation) with running buffer (HBS; 10 mM HEPES, 150 mM NaCl, pH 7.4) in a duplicate dilution series from 3 to 0.1875 μM at a flow rate of 50 μl/min to minimize mass transport effects. Surface regeneration was performed with a one minute injection of 2M $MgCl_2$. The association rate constant (*k*on), dissociation rate constant (*k*off), and equilibrium dissociation constant (*K*d; *K*d = *k*off/*k*on) were calculated assuming a Langmuir 1:1 binding model with BIAevaluation software (GE Healthcare, Uppsala, Sweden). Only data sets with $\chi^2 < 0.5$, were used. During flow through with SPP membrane protein a separate running buffer (50 mM Hepes, 150 mM NaCl, 0.1 % detergent, pH 7.5) was used.

Complex formation between 3D5/EE_48 and three ligands, scFv-EE_3, MBP-EE and SPP-EE, was assessed by SEC as discussed in Chapter 3. Equimolar volumes of purified 3D5/EE_48 and the peptide-tagged protein were combined and allowed to incubate on ice for six min followed by separation using an analytical Superdex 75 gl column (GE Healthcare) equilibrated with 10 mM HEPES, 150 mM NaCl, pH 7.4. Fractions of interest were concentrated and characterized by SDS-PAGE. Control experiments applied the same quantity of each species alone as used in complexation experiments. The Superdex 75 gf column was calibrated using a Low Molecular Mass gel filtration calibration kit (GE Healthcare) supplemented with cross-linked albumin (Sigma). Wildtype SPP was combined with purified 3D5/48 scFv at a ratio of about 2:1

and were allowed to incubate on ice. Equimolar volumes of purified 3D5/EE_48 and SPP-EE with just a slight excess of EE-SPP were combined and allowed to incubate on ice before separation using the analyIn followed by separation using an analytical S300 column. Calibration of the column was performed as in Chapter 3.

4.3.6 Protein co-crystallization

3D5/EE_48 complexed with peptide-tagged proteins were crystallized by the sitting drop vapor diffusion method at 4 °C. Conditions were optimized based on those reported for 3D5/EE_48 [chapter three]. 1-2 μL of protein solution in HBS buffer at 3.8 mg/mL chilled on ice was mixed with 1 μL sample of reservoir solution containing 0.1 M Mes (pH 6.4), 0.1 M magnesium acetate, and 20-24% (w/v) PEG 8000.

4.4 Results & Discussion

4.4.1 Analysis of peptide insertion at the C-terminus

The C-terminus and N-terminus are preferred locations for peptide tag insertion when no structural information is available for the target protein. Placing the short sequence at either the front end or tail end of the target protein involves less risk of destroying the structural integrity or functionality and requires no structural information. Only one terminal peptide insertion was selected for further in depth analysis because it was assumed the presentation of the tag at either end of the protein would display similar characteristics. The C-terminal location was chosen because the vectors were already designed to incorporate the short sequences in this region. In the prior chapter, we discussed the successful kinetic analysis of 3D5/EE_48 for the MBP with a single EE peptide presented at the c-terminal region resulting in a ~767nM affinity. The same

affinity experiment involving a MBP with a sequential c-terminal EE-His_6 tag (MBP-EE-His_6), upon expression, purification and immediate kinetic analysis, showed an improved affinity of 389 nM (χ^2 =0.028). The two-fold increase in affinity observed with the protective His_6 tag on the MBP-EE-his may be the result of the peptide's reduced conformational entropy or extended immunity of the peptide tag to random cleavage. Studies with this protein following a week at 4^{o}C incubation, MBP-EE-His_6 showed no affinity for anti-his variants on the Biacore. After a fourth month incubation at 4^{o}C, complete loss of the EE peptide was observed with direct ELISA. This rapid loss of the His_6 tag, followed by the subsequent loss of the EE tag (389 nM to 767nM) indicates this tag's propensity to cleavage. Therefore protective measures (including a flanking amino acid sequence) must be taken if this location is selected for peptide insertion. The retained affinity of the EE tag suggests that the His_6 tag serves a protective role, prolonging the length of time before c-terminal EE-tag cleavage. Kinetic studies with a sequential c-terminal 10XFlag-His_6 tag showed similar cleavage of the His_6 peptide and retention of the inner 10XFlag peptide (data not shown). In addition to protecting from random protease clipping, the displacement of the peptide from the c-terminal end reduces the conformational entropy. Minimizing the degrees of freedom allows higher probability of binding to occur.

Even with the chaperone's lower affinity to MBP-EE (no His_6), the complex form of MBP-EE and 3D5/EE_48 scFv was successfully isolated using SEC [chapter three]. Current co-crystallization attempts with this complex resulted in the lack of crystal formations from co-crystallization attempts and can be contributed to the high entropy of the peptide tag at the c-terminal region. The high entropy conformation observed at c-terminal appears to interfere with the chaperone's initial recognition for that peptide tag. The flexibility of the peptide can also contribute to a hinge-like conformation resulting in

heterogeneous complexes, which are not beneficial during the generation of a crystal lattice. In order for the peptide-specific chaperone to be utilized in a co-crystallization platform successfully, a tight and rigid complex needs to be created that can create a uniform repetitive unit for lattice formation. This can be established by limiting the number of amino acids located on the tail of the target protein. For this reason only short peptide sequences (6 – 10) amino acids are considered for this platform that contain amino acids that restrict the conformation of the peptide, like the proline in the EE tag. If the flexibility of the peptide at the c- or n-terminal ends can be controlled and rigid complexes can be formed, this would provide a high-throughput system for creating protein crystals. Like purification tags, the peptide tag can be inserted with minimal effort by cloning the protein gene into a vector with the peptide sequence encoded [158, 183]. The ease of use — and the fact that no structural information is required, make this platform and location extremely advantageous for co-crystallization as long as the concerns mentioned above are addressed.

4.4.2 Analysis of peptide insertion into inter-domain linkers

EE peptide insertion in an internal loop requires some structural information about the target protein. However, the benefit of this location is its ability to provide a more rigid conformation of the peptide. Achieving a successfully presented peptide for interaction without impacting the overall structural ability an functionality makes this area of insertion challenging. There are two different types of locations for peptide internalization: interdomain regions or within natural loops (interloop). The degree of protein topology information required for each of these models varies from relatively minimal (interdomain) to essential (interloop). The first region studied was the

internalization of EE peptide tags between two domains of antibodies: scFv and TCRs. The EE peptide was placed prior to the glycine linker ($[G_4S]_4$) that connects the two domains with variations in the flanking residues to examine and determine effective peptide accessibility (Table 4.2). Peptide tags were inserted either directly following the first domain or were flanked with 4-5 additional amino acids prior to the peptide insertion to allow for more flexibility in acceptable conformational presentation. Variants with multiple EE tags were initially designed simply to insure binding and accessibility of the peptide as well as potential applications in protein scaffolds for enzyme assembly.

Expression of each of these variants was similar to the original wildtype from which it was derived with the exception of R30A-EE (Table 4.3). R30A-EE, which did not have flanking residues prior to the peptide tag, resulted in very little monomeric scFv expression following purification via SEC. The additive impacts of the restricted conformation of the peptide tag as a result of the inclusion of the Pro4 and the location of the peptide immediately following the first domain appears to have interfered with proper protein folding. The failure to fold into its correct structure results in the scFv's inability to be transported to the periplasm for extraction using osmotic shock. This suggests that the flanking amino acids help to accommodate the conformation of the peptide tag for retained structural integrity.

Table 4.3 Biophysical Characteristics of EYMPME (EE) Peptide-tagged ligands

	Expression (mg/L culture)	*Melting temp (^oC)*	*% Monomeric Protein*	*Kd (nM), 3D5/EE_48 (χ^2)*
MBP-EE-his	26	63.9 ± 0.5	94	389 (0.028)
DO11	~20	-	97	ND[a]
DO11-v_1	~20	-	99	212 (0.092)
DO11-v_2	~20	-	98	17.1 (0.012)
DO11-e3	~20	-	98	236 (0.023)
14B7	4	69.4 ± 0.4	72	ND[a]
R30A-EE	<0.05	-	30	-
R30A-EE_2	2	51.2 ± 1.6	63	25.5 (0.586)
62-EE_3	2	56.2 ± 0.3	84	29.6 (0.598)

[a]ND = not detected

Kinetic analysis of the expressed proteins indicated that all inserted peptides were accessible for interaction to the chaperone scFv (3D5/EE_48). The variants with multiple peptide tags yielded a higher affinity (17-40nM) which can be attributed to: i) improved probability for accessible binding epitope and ii) additional valency for repetitive binding. Linking sequential peptide tags were generated to confirm that at least one tag is located in a more accessible area for peptide binding. However, the drastic difference in affinities seems to indicate that all the peptides were accessible in the multiple EE-ligands, which results in chaperone interaction with several peptide tags. This leads to a valency issue, where the chaperone acts like a zipper releasing one peptide tag and immediately setting down on the peptide immediately following. The ability to bind both peptides results in a solution that consists of two different complex formations, which produces a heterogeneous solution that is not ideal for crystallization.

While the high affinity of the multiple EE insertion ligands is not preferred for our co-crystallization technology, the application of this peptide insertion for other protein

scaffold applications is still being considered. For crystallization, however, the number of EE sequences must be reduced to a single peptide in order to remove valency issues that can lead to heterogeneous solutions of complexes. Two variants with a single EE-peptide preceded by four random amino acids (D011-e3 and D011-v_1) produced a slightly improved affinity (K_D: 210 – 240nM) than protected EE peptides presented on the c-terminal tail of MBP (K_D: 389nM) (Table 4.4). The significantly lower affinity compared to the multiple EE peptide presenting variants (K_D: 17 – 40nM) indicates that additional location optimization must be addressed further. Further analysis is needed to explore if the single EE peptide variants can reach the same affinity as those with multiple EE tags by using the second peptide location. Further location optimize can determine if this location is an improved location or merely an artifact of the valency addressed above.

Table 4.4 Kinetics of 3D5/EE_48 scFv binding to EE-tagged proteins by SPR

ligand	k_{on} (1/Ms)	SD (%)	k_{off} (1/s)	SD (%)	K_d (nM)	χ^2
MBP-EE-his	3.5 E+04	47	1.2 E-02	2	389	0.028
DO11-e3	4.1 E+05	41	8.08 E-02	16	236	0.023
DO11-v_1	2.4 E+05	32	4.77 E-02	21	212	0.010
DO11-v_2	2.7 E+05	41	3.81 E-03	29	17.1	0.012
R30A-EE_2 (frozen)	2.0 E+05	73	3.69 E-03	58	28.9	0.314
62-EE_3 (frozen)	3.6 E+05	63	8.17 E-03	26	40.2	0.131
R30A-EE_2	3.9 E+05	30	9.0 E-03	34	25.5	0.586
62-EE_3	3.6 E+05	37	9.2 E-03	41	29.6	0.598

4.4.3 Analysis of peptide insertion into internal loops: within natural loop

The second region for internalized peptide insertion that we focused on was a natural loop of maltose binding protein. Unlike the scFvs and TCRs, where the peptide insertions are guided by knowledge of artificial linkers locations, selection of the peptide insertion region requires a little more finesse and more detailed structural information. Utilizing the MBP PDB (#1YTV and 1A7L) surface accessible natural interloop regions

with (~6-8 amino acids in length) were examined. Careful consideration needed to be taken when choosing the kinked loop due to Pro4, which limits the conformational entropy. Amino acids 171 – 176 (sequence KYENGKY) were selected as the ideal region due to its conformation and inclusion of amino acids with similar properties to those in the EYMPME sequence (Fig. 4.1a). The initial variant (MBP-v_1) was designed with flanking linkers to provide steric accessibility of the peptide. Additional variants were modified to reduce the number of flanking linkers and to determine if direct replacement of the loop was acceptable (MBP-v_2 and MBP-v_3).

Table 4.5 Biophysical Characteristics of Peptide-tagged MBP variants

	Expression (mg/L culture)	*Melting temp (^oC)*	*% Monomeric Protein*	*Kd (nM), 3D5/EE_48 (χ^2)*
MBP-his	24	63.3 ± 0.3	99	ND[a]
MBP-EE	8	62.5 ± 0.2	N/A	767 (0.030)
MBP-EE-his	26	63.9 ± 0.5	94	389 (0.028)
MBP-v_1	14	-	94	NSB[b]
MBP-v_2	32	-	98	1210 (0.224)
MBP-v_3	18	52.8 ± 0.8	96	816 (0.024)

[a]ND = no detection binding
[b]NSB = non specific binding

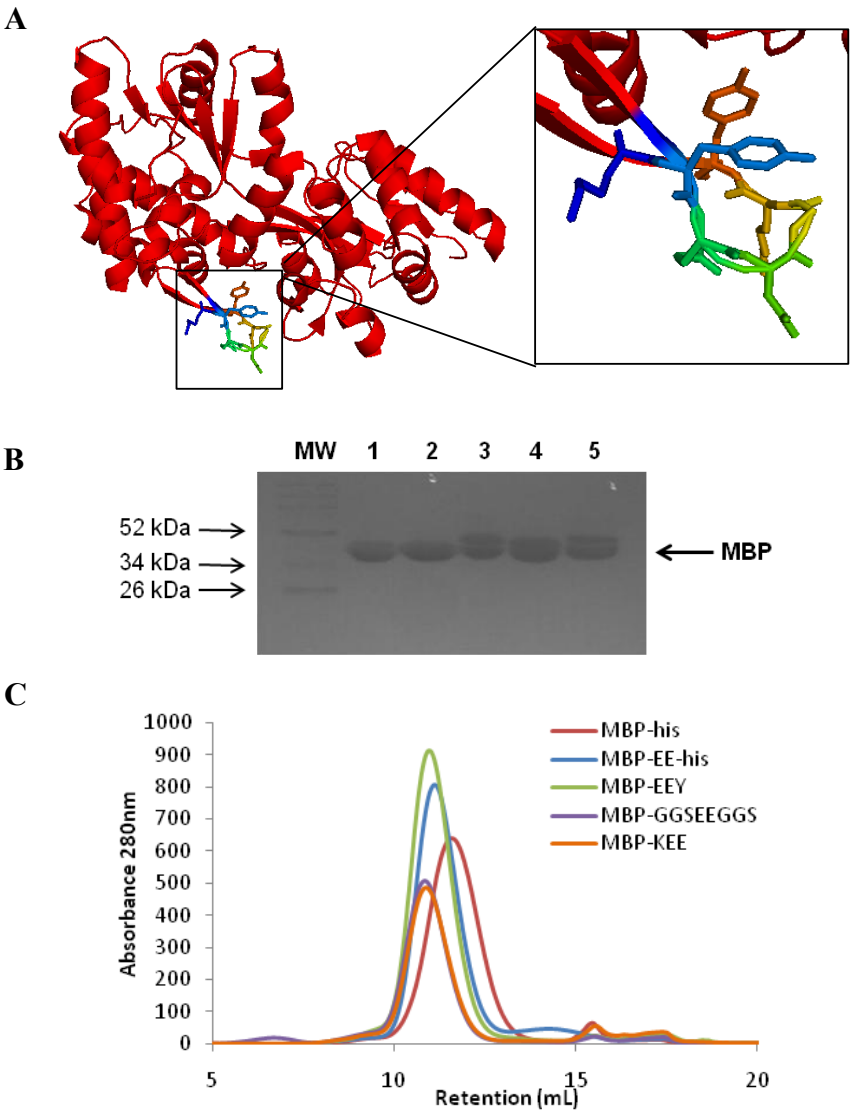


Figure 4.1 Characterization of the MBP peptide presenting variants

(a) Structural Image of the MBP natural loop (171-176: KYENGKY,) selected for peptide insertion with an enlargement of the loop configuration with peptide rainbow color coded from n-terminal blue to c-terminal red.

(b) Size and purity by SDS-PAGE: Lane 1: MBP-his; Lane 2: MBP-EE-his; Lane 3: MBP-v_2; Lane 4: MBP-v_3; Lane 5: MBP-v_1.

(c) Purification by SEC. Molecular weight standards on S75 elute at 7.4 ml (void volume; blue dextran 2000); 9.6 ml (75 kDa conalbumin); 10.9 ml (43 kDa ovalbumin); 12.7 ml (29 kDa carbonic anhydrase); 15.4 ml (residual peak from Ni^{2+} affinity column).

Kinetic analysis, using Surface Plasmon Resonance (SPR), established that the peptide with the flanking linkers (three amino acids on each side) resulted in non-specific binding of the MBP (not shown). The addition of six flanking amino acids may have affected the structural conformation, resulting in improper folding of hydrophobic regions. The binding curves for the maltose binding proteins variants were very similar in nature, MBP-v_2 and MBP-v_3, however the resulting affinities were lower than those for the EE c-terminally tagged maltose binding proteins (Fig. 4.2 and Table 4.5). This shows that minor modifications to the loop, where the peptide is directly substitutes the natural loop, still allow for successful affinity to the EE-specific scFv chaperone. The successful kinetic analysis of the direct replacement of the loop indicates that peptide insertion in either format does provide a somewhat accessible peptide although some structure stability is lost. The drastic change in MBP-v_3's melting temperature to 52.8 ^{o}C (± 0.76), from the wildtype Tm of 63.9 ^{o}C (± 0.46), indicates the altered stability of this variant and suggests that there was a perturbation in the structural integrity of the maltose binding protein (Table 4.5). The visualization of doublets from SDS-page gel separation across all MBP variants appears to confirm the possibility that the structural stability of the all the variants were affected during peptide insertion (Fig. 4.1).

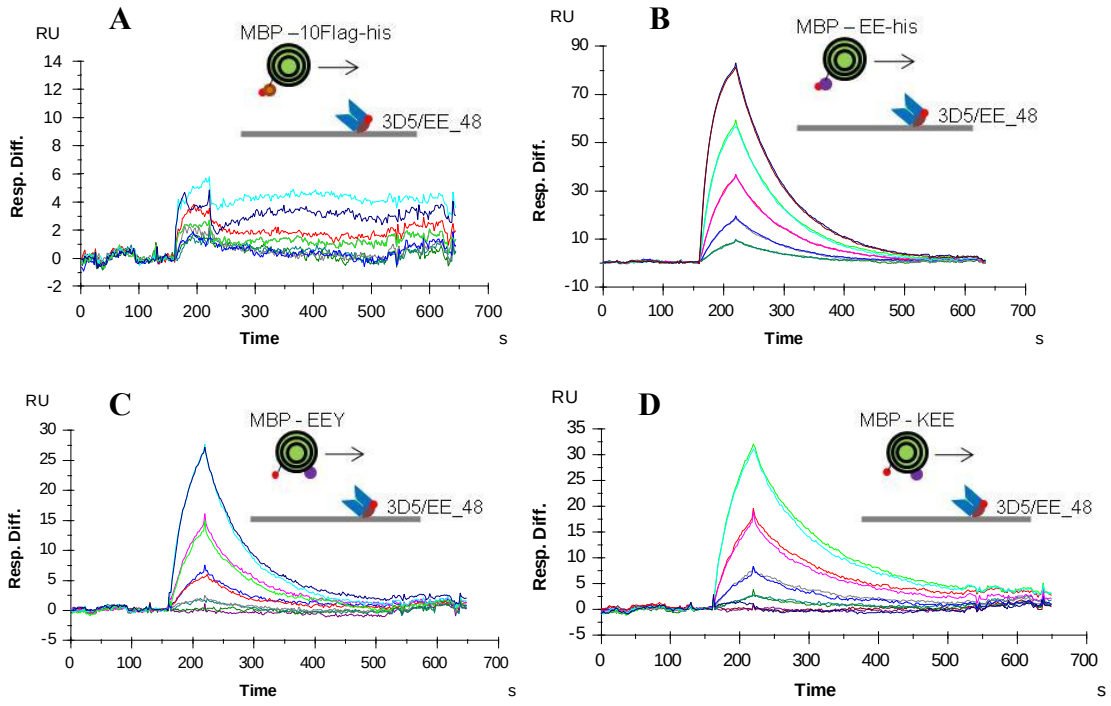


Figure 4.2 SPR analysis of comparison of different variants of MBP with EE peptide insertion at internal loop (171 – 176).

Coupled on the chip is chaperone 3D5/EE_48 scFv chaperone with wildtype14B7 scFv couples as the blank. Serial 1:2 Dilutions of soluble proteins were injected over the chip from 3uM to 0.1875uM

(a) MBP with c-terminal Flag-His_6 – negative control
(b) MBP with c-terminal EE-His_6
(c) MBP-v_2: EYMPMEY
(d) MBP-v_3:KEYMPME

Table 4.6 Characterization of 3D5/EE_48 scFv binding kinetics with internal loop peptide presenting proteins by SPR

Ligand	k_{on} (1/Ms)	SD (%)	k_{off} (1/s)	SD (%)	K_d (nM)	χ^2
MBP-EE	1.5 E+05	62	7.3 E-02	24	767	0.030
MBP-EE-his*	3.5 E+04	47	1.2 E-02	2	389	0.028
MBP-v_2**	1.4 E+04	44	1.5 E-02	10	1210	0.024
MBP-v_3	1.2 E+-4	29	9.1 E-03	33	816	0.025
SPP-EE	3.2 E+04	57	3.1 E-02	12	1283	0.024

*Highest affinity
**Lowest affinity

For internalization of the tag at the 171-176 natural loop in MBP, the six flanking residues were not preferred. Direct peptide substitution resulted in chaperone affinities that were two-fold lower (K_D~800nM) than those for the protected c-terminal tag (K_D: 387nM) and even the unprotected c-terminal tag (K_D: 767nM). The low affinities of the MBP EE-tagged variants highlight the challenges involved with optimizing the peptide insertion in this natural loop. The addition of extra amino acids should allow for more flexibility in this loop, resulting in improved accessibility of the EE tag. However, the flanking triplet residues in MBP-v_1 may have created too much structural perturbation on the natural loop conformation. The current internalized peptide need to be further optimized for accessibility and rigidity to establish a favored conformation. Additional variants with varying flanking residues are needed to study the extent of steric accessibility these residues can provide without introducing added conformational entropy or upsetting natural protein folding. It should be considered that a minimal flanking linker is preferred for this peptide tag because of its tendency to increase the rigidity of the resulting complex thereby enhancing its propensity to crystallize.

4.4.4 Application with a membrane protein

The inherent flexibility of many intracellular loops on membrane proteins can limit the crystal formation and even the quality of the crystals [74]. The design of this peptide-specific chaperone technology aims to specifically target and replace flexible unstable regions with the peptide tag, allowing the chaperone protein to further lock these regions in place. Although structural information is very limited for most membrane proteins, solvent-exposed loops have been identified through other sources such as homology sequencing analysis or membrane topology studies [184]. The Kobilka group

successfully used these methods to systematically remove flexible regions and even insert an entire T4 Lysozyme in place of one loop [61, 93, 94]. If a solvent-exposed loop can be identified, introduction of a peptide in this region can be done with relative ease. The resulting complex of chaperone with the peptide would rigidify this otherwise flexible region and remove a critical region that would previously interfered with crystallization.

To confirm the ability to apply our platform for membrane proteins, the EE peptide tag was inserted into a flexible extra-membranous loop and terminal regions of the highly conserved transmembrane enzyme, signal peptide peptidase (SPP). The complex formation, between our chaperone and the SPP with an internal EE peptide, simultaneously immobilizes the SPP's flexible loop while providing additional hydrophilic surfaces for lattice formation. One particular variant with a single EE tag located at an extra-membranous loop produced an affinity for the 3D5/EE_48 chaperone of 1283nM (χ^2=0.024) (Fig. 4.3a), which was similar to the MBP-v_2 variant (K_D: 1210nM). A direct kinetic analysis comparison of these two variants is difficult due to the high noise signal contributed by the surfactant buffer that stabilized the membrane protein. The 0.1% detergent in the SPP running buffer may be negatively impacting the binding interactions.

Regardless of the low uM affinity, the peptide was accessible enough for complexation with our 3D5/EE_48 chaperone (Fig. 4.3c and Fig. 4.3d). While the chromatograms do not show much difference between the control and the EE-tagged SPP, the separation of column flow fractions over SDS-page gels display successful complexation of the SPP-EE with our hyper-crystallizable chaperone 3D5/EE_48 scFv (Fig. 4.3c and Fig. 4.3d). The initial peak in the SPP-EE SEC flowthrough now contains both SPP-EE and the scFv with the leftover uncomplexed scFv present at the tail-end of the separation. Successful isolation of the complexed chaperone with SPP proved a little

more difficult as a result of the surfactants in the running buffer. A major concern now is to efficiently separate the coupled SPP from non-complexed SPP, as the two peaks nearly overlap. Column resolution using gel filtration has always been a challenge in membrane protein fractionation; however, other purification methods can be utilized to successfully separate the complexed form from the uncomplexed. Current work is in progress to utilize Ni columns, following size exclusion to get the correct complex for co-crystallization trials. In addition, current SPP variants now contain two EE tags, one at an internal loop, and the other at the n-terminal region of the protein. The design of this protein will further show the utility of our chaperone's ability to successfully bind the peptide at different locations. The objective of utilizing multiple peptide locations is to simultaneously surround the more hydrophobic regions of the protein with additional polar regions to mediate contacts. Current co-crystallization trials are in progress for complexed chaperone with the model membrane protein, as well as many of the EE-tagged proteins mentioned above. Successful complex crystallization will allow us to ultimately determine if these peptide insertions are acceptable for our co-crystallization platform.

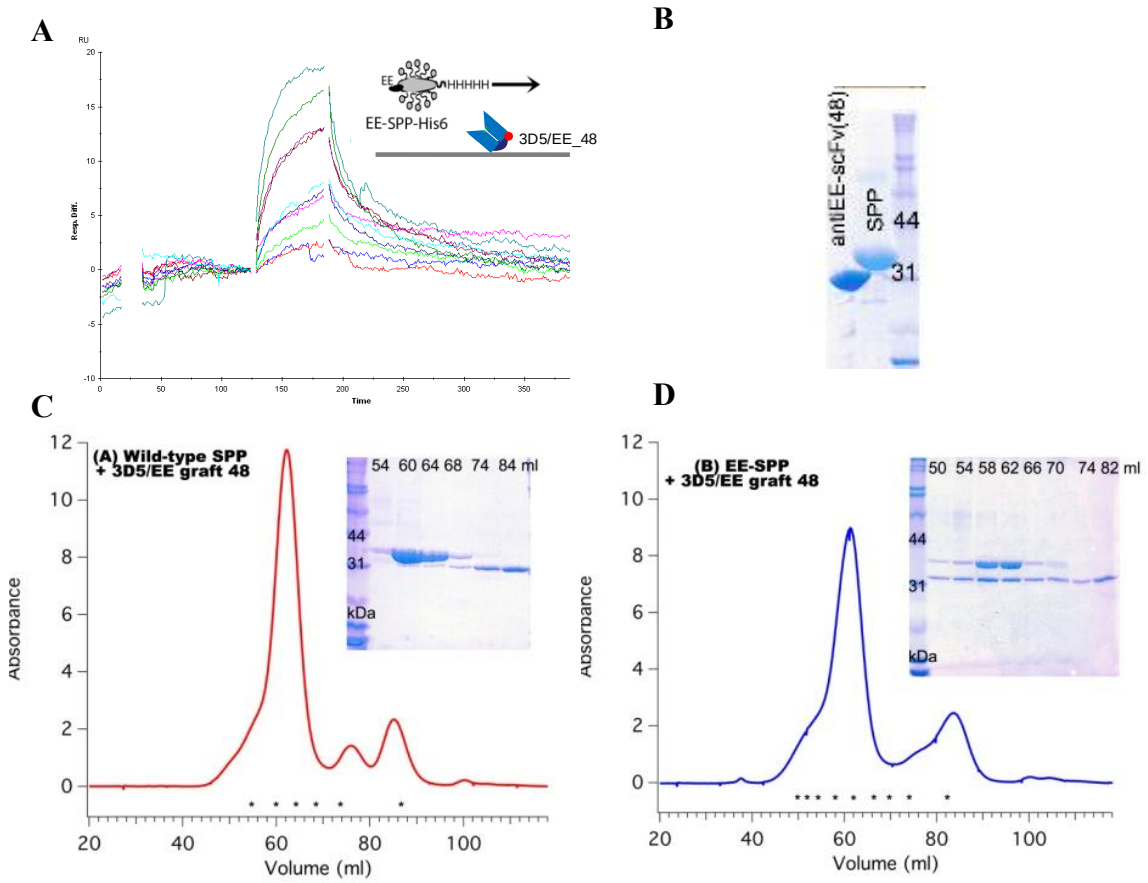


Figure 4.3 Characterization of the chaperone and SPP-EE interaction.

(a) Preliminary kinetic analysis of the SPP-EE protein (internal loop) Serial 1:2 Dilutions of sample were injected over the chip from 3uM to 0.1875uM with 14B7 scFv coupled as the blank

(b) Size and purity by 12% SDS-PAGE

(c) Isolation of 3D5/EE_48 complexes with variants SPP by SEC: 3D5/EE_48 incubated with wild-type SPP

(d) Isolation of 3D5/EE_48 complexes with variants SPP by SEC: 3D5/EE_48 incubated with SPP-EE. For (c) and (d), SDS-PAGE analysis of elution peaks (top right of each respective image) confirms protein and complex identities. The three peaks visualized on the chromatograph is first the SPP, followed by the dimeric and monomeric forms of the chaperone scFv.

4.5 CONCLUSION: FUTURE POTENTIAL USE FOR CO-CRYSTALLIZATION CHAPERONES

Peptides were inserted at three locations: terminally, interdomain and interloop. All peptide-tagged proteins have shown successful complexation with the 3D5/EE_48 chaperone even at low affinity (K_D~1uM). Terminal peptide insertions are the preferred insertion site if no structural information is required. For internal peptide insertions, consideration for the EE tag's restricted conformation required us to address the use of additional linkers that provide flexibility and accessibility for the tag. A single EE tag insertion with 4-5 flanking residues prior proved beneficial for insertion in our interdomain proteins to create rigid complexes and create a homogenous solution for crystallization. The peptide insertion into the natural loop of MBP and SPP shows the difficulties of selecting an acceptable loop and retaining a good peptide conformation. The guidelines we have established here for peptide insertion are bringing us a step closer to creating the generalized toolbox of co-crystallization peptide-specific chaperones. Observations of these EE-tagged protein variants will provide additional information as we generate guidelines for preferred insertion that can be successfully transferred for use in other target proteins. The ability to insert the EE tag into different regions of the protein and retain successful chaperone complexation shows the broad application of system. This technology is potentially transformative, allowing the incorporation of the peptide into any membrane protein loops/tails and no longer requiring the need to re-engineer chaperones for each new membrane protein target.

CHAPTER FIVE: Control of Protein Crystallization through Tuning Intermolecular Contacts

5.1 CHAPTER SUMMARY

The arrival of protein engineering techniques for improving crystallization, such as SER (surface entropy reduction) and chaperone assisted co-crystallization, have indicated that a protein's intermolecular contacts assist in drawing proteins into an ordered lattice. Analysis within a selection of scFv antibodies (14B7 and 3D5 derived variants) revealed that modulating an intermolecular contact's energetic contribution can guide the crystal lattice formation for the protein. Critical residues involved in a particular intermolecular contact can be determined using *PISA* and *Rosetta Alanine Scanning*, where the criteria for high affinity interactions can similarly be applied to low affinity crystal contact interactions. Global manipulation of 3D5 scFv's surface epitopes affected the protein's intermolecular contacts resulting in a wide range of crystal lattice formations (hexagonal, tetragonal and cubic). Armed with current protein engineering strategies for improved crystallization, we can systematically modulate the energetics of the 3D5 based antibodies' topology for improved hyper-crystallizability. The creation of co-crystallizable chaperones will potentially transform traditional shotgun crystallization screening techniques into a methodology with a more rational approach and higher-success rate.

5.2 INTRODUCTION

Co-crystallization is a method that utilizes a chaperone protein to assist the crystallization of a target protein. Frequently used for membrane protein like GPCRs

[61, 93, 94], the addition of the second protein assists in crystallization by protecting hydrophobic regions, replacing or locking in flexible regions and most importantly providing additional polar regions for more potential lattice contacts [79, 85, 88, 92, 98, 179]. Currently a universal hyper-crystallizable antibody chaperone is in development that is specific to a peptide tag [Chapter 3]. This peptide tag can be appended to any target protein of interest and will assist in drawing the intended target protein into a crystalline structure. [Chapter 4]. With this platform technology, crystal optimization efforts now focus on a well characterized peptide specific chaperone with improved biophysical characteristics and higher propensity for crystallization. Emphasis is placed on engineering a chaperone with the ability to mediate crystal contacts utilizing many of the protein engineering modification techniques in place.

Crystallization requires molecules to associate together, mediated by weak specific interactions. Surveys of experimental results from the Northeast Structural Genomics Consortium, has led to the conclusion that surface epitopes drive the propensity for crystallization by providing intermolecular interactions [37]. Stringent statistical analysis indicates that the intermolecular contacts are generated by anisotropic interactions that include small hydrophobic residues and exclude high-entropy residues with large polar side chains [185, 186]. The many interaction contacts available for hyper-crystallizable lysozyme's different lattice formations appear stochastic in nature [64]. Molecular dynamic simulations of hyper-crystallizable lysozyme, however, indicate that the extent of these interactions is dependent on the physical chemistry involved at the interface [186]. This suggests that the nucleation of lysozyme into its multitude of space groups may be initiated in a systematic fashion contrary to previous perceptions that crystallization was generated largely by random interactions [17, 187, 188].

The rise of new recombinant technology produced a drastic increase in material for protein crystallization by reducing protein purification efforts with fusion proteins and affinity tags [183, 189, 190]. Crystallization is no longer dependant on samples of purified native protein, however the preparation of high-diffraction quality crystals still remains the limiting factor to structural determination. Current strategies for crystallization primarily focus on extensive shotgun-screening of thousands of crystallization conditions, which is a tedious and time-consuming process [191]. While these strategies focus on improving the conditions of crystallization, recent progress has emphasized the need to modify the protein molecules themselves to enhance crystallizability [10]. Progress in protein engineering has improved the chances of creating target protein crystals by modifying biophysical characteristics of proteins [37, 78, 177, 192]. These modifications allow successful crystallization to proceed, which had not been attained with the wildtype form. In addition, proteins that crystallized poorly can be modified using these techniques for higher diffraction quality crystals. Protein modification techniques such as solubility [149] or surface entropy [37, 74, 185, 193] impact the intermolecular contacts involved in crystallization. Improved solubility allows for more stable proteins less prone to the random interactions characteristic of aggregation [57, 59]. The generation of more symmetric proteins allow for well-ordered packing to proceed. Surface Entropy Reduction (SER) and flexible protein truncation both aim to accomplish well-ordered packing by reducing the entropic cost of the protein. Truncating flexible regions of the protein can eliminate their propensity to create multiple interaction contacts [193] and reduce the conformational heterogeneity, characteristically seen in many membrane proteins, resulting in improved crystals with higher resolution [74]. SER is usually intended for proteins that are intrinsically stable, targeting specific flexible amino acids (Glu, Gln, Lys) [37, 185] and replacing them with threonines and

tyrosines which are known to contribute to protein-protein interfaces [87] or alanines, a minimal residue [68].

Utilizing the protein engineering modification techniques in place, the improved biophysical characteristics mediate novel crystal contacts or improve the preexisting protein-protein interactions. Surface modification of a potential antibody chaperone for co-crystallization produced novel intermolecular contacts that impact the propensity for new crystal lattice frameworks [Chapter 3]. These crystal contacts are viewed as low affinity interactions and their energetic must be studied in order to better understand their role in guiding crystallization. Understanding the energetics of intermolecular contacts and their ability to define lattice formation will allow us to use current surface modification techniques (SER, optimization of target constructs) and potentially provide additional considerations to improve the antibody chaperone's crystallizability. Two different murine model family antibodies will be studied to examine the global effects modulating crystal contact energetics has on lattice formation in crystallization (anti-hexahistidine antibody 3D5 scFv, [99]). Computational modeling program and crystallization experiments were used to investigate whether protein crystallization and lattice formation can be controlled by fine-tuning the energetics of key intermolecular contacts. The impact of manipulating a particular intermolecular contact on crystal formation (anti-protective antigen of anthrax antibody, 14B7 scFv [39, 41]) will potentially allow us to better govern crystallization of our antibody chaperone.

The results presented here highlight the potential role intermolecular contacts play in the two-step process of crystallization: first clustering of solute molecules followed by reorganization into ordered nuclei is of primary interest. Our understanding of these interprotein interactions and their role in crystal contact formation will allow for rational modification of our universal chaperone for enhanced crystallizability.

5.3 MATERIALS AND METHODS

5.3.1 Computational analysis of amino acids selected for mutation

Experimental structures derived from diffraction data found in the protein data bank were used in computational analysis (3D5 family: PDB #1KTR, 3NN8; 14B7 family: PDB #3ESU, 3ETB). *PDBe Protein Interfaces, Surfaces and Assemblies (*PISA) was used exclusively to examine the critical amino acids in the crystal contacts that compose the 3D5/EE_48 lattice framework. Amino acids were selected on based on two criteria: i) critical H-bond forming amino acids and ii) potential salt bridges (charged amino acids). One final amino acid was selected based on the determination of its role in the original 3D5 crystal lattice as a potential electrostatic interaction. *Rosetta's Computational Mutagenesis Alanine Scanning* program was also used as an additional selection tool to predict energetically important amino acid in each intermolecular contact. The program functions by replacing each amino acid at the crystal contact with an alanine and predicting the residues with the highest loss of free energy ($\Delta\Delta G \geq 1.0$ kcal/mol) with respect to the alanine substitution [194, 195]. The Rosetta and PISA programs were used in combination to study the energetics involved with protein-protein interaction for the 14B7 family of murine antibodies. The areaimol program from the *Collaborative Computational Project (CCP4) software for Macromolecular X-Ray Crystallography* provided the calculations for the solvent accessible surface areas of the residues as well as the surface area involved in the crystal contact [196]. Average $\Delta\Delta G$s were calculated for the 14B7 intermolecular contact using *Alanine Scanning Robetta* software by averaging the results from the two separate occurrences of this interaction in the asymmetric unit. Amino acids were selected that showed altered free energy

contribution to the complex on alanines substitution based on the Robetta software and contribute a significant portion of the surface area without impacting the protein's stability. Amino acids identified as critical for loop conformation in literature [197, 198] were avoided with the exception of residue L:W33, which was included because the amino acid exposed area involved in the crystal contact was greater than 85%.

5.3.2 Molecular biology of scFv proteins

Each of the scFv variants was created by linking a V_L and V_H chain with a gly-ser linker [$(G_4S)_4$]. Targeted amino acids, selected from computational analysis, were mutated to preferred amino acids in proteins via site-directed mutagenesis with mutagenic oligonucleotides (Integrated DNA Technologies). For 14B7 scFv, the amino acids with $\Delta\Delta G \neq 0$ kcal/mol were mutated to alanines (Appendix B.8.1). Double alanines substitutions were selected with one alanines mutation in each chain (one heavy and one light) and introduced using two sequential site directed mutagenesis steps (Appendix B.8.1). The amino acids selected in 3D5/EE_48 scFv were mutated to the original wildtype scFv (3D5) amino acid or to an alanine if that location was not originally present in the wildetype scFv. Five variants were engineered to target different two different crystal contacts in the 3D5/EE_48 lattice formation: i) 3D5/EE_48.2 – S30T and S32T ii) 3D5/EE_48.2a – S30T and S32A iii) 3D5/EE_48.3 – D55G, K57T, R58S and S62K iv) 3D5/EE_48.23 – combined substitutions of (i) and (iii) v) 3D5/EE_48.23a – combined substitutions of (ii) and (iii). Variants utilizing the anti-hexahistidine 3D5 scFv framework [99] were derived by generating libraries with site-directed CDR mutagenesis, selected and screened as previously mentioned [Chapter 2 and 3].

5.3.3 Protein expression, purification and characterization

All engineered proteins were subcloned into the *Sfi*I-*Sfi*I site of pAK400 for scFv expression [158] and contain a c-terminal, vector encoded His_6 tag to facilitate purification. Protein was secreted into the periplasm of *E. coli* strain BL21, by induction of 1mM IPTG (OD600 = 3.0) for 5 hours at room temperature. The protein was isolated by osmotic shock and purified using immobilized Ni^{2+} affinity affinity chromatography (IMAC) followed by size exclusion chromatography (SEC) using a Superdex S75 column (GE Healthcare) equilibrated with HBS running buffer (10 mM HEPES, 150 mM NaCl, pH 7.4), as previously described [160]. The Superdex S75 column was calibrated using a Low Molecular Mass gel filtration calibration kit (GE Healthcare). Qualitative analysis of the distribution of multimers at equilibrium in solution (scFv monomer-to-multimer ratio) was determined from SEC traces by calculating the area under the curve for each peak using Unicorn software (GE Healthcare).

Protein purity and size were characterized by 12% SDS-PAGE under reducing conditions [161]. Protein concentration was assessed by BCA assay with a BSA standard curve and buffer blank (Pierce). Protein solubility was determined by quantifying the concentration of the remaining soluble protein solution following a concentration of the protein to ~20 mg/mL and incubation for four days at 4 °C with centrifugation (10min at 13.2krpm) to pellet insoluble particles. Maintained structural conformation for each mutant was analyzed using two methods: circular dichroism and thermal stability with Real Time PCR (RT-PCR). Circular Dichroism (CD) measurements were performed with secondary structure analysis using the yang reference data set [199]. Each protein's structural stability was assessed as the mid-point for complete thermal unfolding, using a

fluorescence assay [162]. Purified protein (20 µl at 280 µg/mL) or buffer blank were mixed with Sypro Orange (1 µl of a 1:1000 dilution; Molecular Probes) and heated in a RT-PCR instrument (7900HT Fast Real-Time PCR System (Applied Biosystems) at increments of 0.5 °C from 20 – 85 °C and analyzed with SDS.2 (Applied Biosciences). As we mentioned in chapter three, to facilitate direct comparisons, all 3D5 and variant characterization values reported here were performed with these methods and specific values may differ slightly from those previously reported [149].

Impact of the 3D5/EE_48 amino acid substitutions on its affinity for the EE tag was established through kinetic analysis with Surface Plasmon Resonance (SPR). Kinetic binding assays were performed with a MBP bearing a c-terminal EE tag protected by a His_6 tag to quantify the 3D5/EE_48 selectivity using a BIAcore 3000 (GE Healthcare). Protein ligands were coupled to CM5 chips using NHS-EDC chemistry to a level of ~500 RU. Responses due to sample refractive index changes were corrected using signal from a flow cell coupled with a control protein (MBP bearing a c-terminal 10XFlag-His_6). Dilution series from 3 to 0.1875 µM of soluble proteins was injected in duplicate at a flow rate of 50 µl/min to minimize mass transport effects. Surface regeneration was performed after each run with a one minute injection of 2M $MgCl_2$.

5.3.4 Protein Crystallization

All 14B7 scFv variants were crystallized via sitting drop vapor diffusion under the same conditions as the wildtype (20% PEG 4000, 0.1M HEPES (~ pH 7.5–8.5), 10% IPA) at 20^{o}C for up to 6 months. Crystal trials were prepared with filtered proteins (0.25um) concentrated to 14mg/mL using 1uL sitting drops at varying concentrations of precipitant to protein (protein to mother liquor ratios of 1:2, 1:1 and 2:1). Proteins

crystals were grown to a minimum size of 20 μm and preliminary confirmed using polarized light. Crystallization rates were determined by studying the crystal size every three hours. 3D5/EE_48 and its variants were crystallized as previously discussed in chapter 3. Crystals of rectangular or triangular shape grew within 4 weeks to a maximal size of 40-60 μm. 3D5/His_683 was crystallized by optimizing the conditions of 3D5/EE_48, refer to Table 5.1 for the various conditions tested. Crystals of teardrop or rodlike shape grew within 4 days. Crystallization of 3D5 wildtype scFv has not been successfully replicated in our lab and therefore all mention of its crystal contacts is in reference to published literature [99].

Table 5.1 3D5/His_683 Crystallization Conditions at 4^{o}C

Protein Conc. (mg/mL)	ML:Protein Ratio (uL)	Mother Liquor Component 1	Mother Liquor Component 2	Mother Liquor Component 3
10.0	1:1	0.1M Imidazole (pH 8.0)	0.1M LI_2/Na_2SO4	17-22% PEG3350/4000/8000
3.8	1:1 w/His_6	0.1M Imidazole (pH 8.0)	0.1-0.2M LI_2SO4	11-16.4% PEG 3350
3.8	4:3, 1:1 2:1, 5:3	0.1M MES (pH 6.4)	0.1$Mg(OAc)_2$	20-25% PEG4000
3.8	1:1	0.1M MES (pH 6.4)	0.1-0.25M $Mg(OAc)_2$	17.5-30% PEG8000
4 - 10	1:1	0.1M Imidazole (pH 8.0)	0.4 Na_2HPO_4/ 1.6 KH_2PO_4	0.25M NaCl

5.3.5 Data collection, structure determination and refinement

14B7variant crystals were transferred into a cryoprotectant solution consisting of the reservoir solution and 25% (v/v) glycerol. Each crystal was then mounted in a cryoloop (Hampton Research, Laguna Niguel, CA) and dipped into liquid nitrogen for placement in the coldstream. 14B7 crystallographic data was collected at the University of Texas at Austin's beamline. Structure determination and refinement were performed

using diffraction methods imosflm [200] and Molrep [201] from the *CCP4 software* [164]. 3D5 family scFv crystals were harvested at 25°C and cryocooled using a solution consisting of 85.5% (v/v) reservoir solution and 14.5% (v/v) ethylene glycol. 3D5 family scFv crystallographic data were collected using a wavelengths of 1-6 Å at the GM/CA-CAT beamline (Darien, IL) equipped with a 5 μm mini-beam setup. Data were processed with XDS [163] and Scala [201]. The structures were solved by molecular replacement with Molrep using a polyalanine search model derived from parent 3D5 asymmetric unit (PDB ID 1KTR) from which all non-protein atoms and loop residues were removed. Translation/Libration/Screw motions and medium non-crystallographic restraints was conducted. Figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 0.99rc6, Schrödinger, LLC).

5.4 RESULTS AND DISCUSSION

5.4.1 The role of energetics in crystal contacts: critical amino acid residues

The intermolecular contacts involved in crystallization are characterized as low affinity interactions. While it has been shown that surface epitope manipulations can impact crystallization [57, 68, 177], it has not been established if these low affinity interactions can be controlled on the residue level as high affinity interactions are [112, 194, 195]. Systematic mutations at the crystal contact region will determine if modulating the energetic contribution of an protein-protein interaction can guide the formation of crystal lattice contacts. A model scFv with a large interprotein interaction (789.3 Å) was used to study the intermolecular contacts role in driving crystallization. The 14B7 wildtype (WT) scFv antibody was selected because this large intermolecular contact was retained even with 10 mutations (Light: L46F, S56P, S76I, Q78L, L94P;

GlinkD; Heavy: S30N, T57S, K64E, T68I) and a change in space groups (M18 scFv, Table 5.2). Both scFv antibodies were crystallized in the same conditions 20% PEG 4000, 0.1M HEPES (~ pH 7.5–8.5), 10% IPA, however 14B7 crystallized within 24 hours into a primitive tetragonal space group lattice structure ($P42_12_1$) [39] while the ultra-high affinity 14B7 variant with 10 amino acid changes (M18) required 6 months to form crystals and resulted in a P21 space group [41] (Fig. 5.1a and Fig. 5.1b). Structural data for 14B7 wildtype and the affinity matured scFv M18, were used to determine low affinity protein-protein interactions (Table 5.3).

Table 5.2 14B7 Survey of Crystal Contacts in the Unit Lattice Cell

Contact #	Chains involved	Interface area ($Å^2$)[a]
1	V_L+V_H:V_L+V_H	789.3
2	V_L:V_H	436.7
3	V_H:V_H	262.7
4	V_L:V_H	158.9
5	V_H:V_H	98.8

[a]Interface area based off of PISA

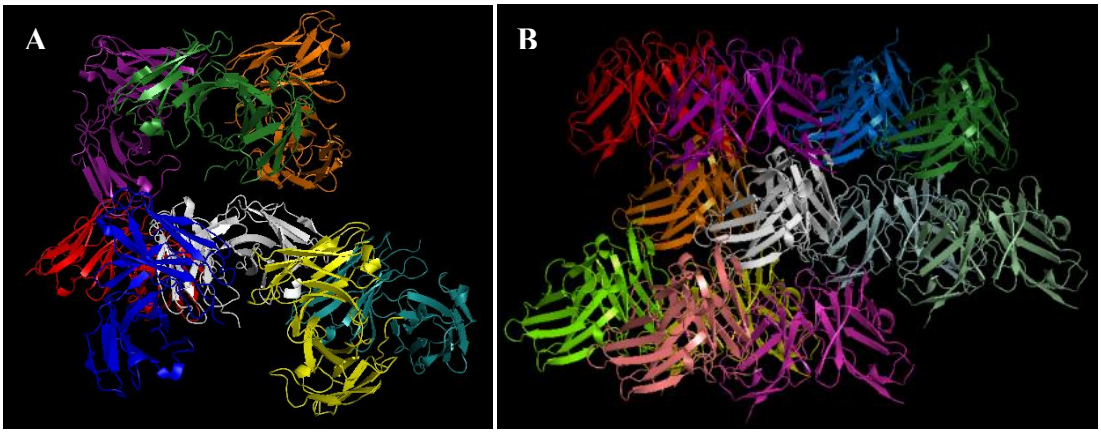


Figure 5.1. Crystalline Lattice Structure of 14B7 variants

(a) 14B7 (Space Group: $P42_12_1$);
(b) 14B7 variant M18, with 10 mutations (Space Group: P1)

Within the lattice unit cell for the wildtype 14B7, there are eight scFv molecules with 13 contact points, of which five are distinct low affinity protein-protein interactions (within 7Å) characteristic of a crystal contact (Fig. 5.1a). The largest interaction has a protein-protein contact, nearly two times larger than any other contact (Table 5.2). Its retention in the M18 variant crystal lattice (primitive P1) in spite of the different growth kinetics [41], emphasizes the need to study its low affinity energetic surface epitopes in more detail. By modulating this specific contact, it will allow us to study the energetics involved in the driving crystallization. Critical intermolecular amino acids were selected using PISA, which highlights predicted residues that contribute hydrogen bonds and electrostatic interactions (salt bridges). Rosetta's Computational Mutagenesis Alanine Scanning program [194, 195] was used to select amino acids predicted as energetically important to the major protein-protein interaction ($\Delta\Delta G \geq 1.0$ kcal/mol, after substitution with alanines) (Table 5.3). Although Rosetta's alanine scanning is commonly used to study the energetics involved with high affinity binding, it is believed the energetics

involved are similar to those in low affinity interactions like crystal contacts and can therefore be applied for crystal contact determination. It is unclear if the 1 kcal/mol threshold is too high for analysis of crystal contacts, considering that the strength of binding interactions are greater than intermolecular contacts involved in the crystal lattice.

Table 5.3 Characterization of Critical Residues in 14B7 Major Crystal Lattice Contact

Chain: Kabat#	Amino Acid (AA)	Δ ΔG in 14B7 (kcal/mol)[a]	Δ ΔG in M18 (kcal./mol)[a]	Hydrogen bond Electrostatic Interaction[b]	SA in 14B7 Crystal Contact (Å^2)[c]	% SA of AA in 14B7 Crystal Contact	% of 14B7 Total Crystal Contact
L:28	D	-0.20	0.94	-	52.3	64.7	7.4
L:30	R	2.66	0.26	HS	130.5	82.9	18.6
L:32	Y	2.17	0.34	H	69.7	94.3	9.9
L:50	Y	2.06	0.03	H	69.0	90.7	9.8
L:53	R	0.22	0.53	-	27.7	22.7	3.9
H:31	S	0.14	-0.17	H	46.7	66.7	6.6
H:33	W	0.87	-	H	15.5	89.6	2.2
H:54	D	0.65	-0.13	HS	85.0	75.5	12.1
H:56	D	1.19	-0.34	HS	36.3	49.1	5.2
H:97	L	1.16/0.66	1.54	-	131.6	81.6	18.7
H:98	L	2.18	1.29	-	38.9	76.3	5.5
H:99	R	0.91	-	H	62.6	46.8	8.2

[a] Computational mutagenesis Alanine Scanning Robetta Software (average of all four antibodies involved in the two disctinct contacts)
[b] Predicted using Protein Interfaces, Surfaces and Assemblies (PISA), H- hydrogen bond contribution, S- Salt Bridge contribution
[c] SA=Surface Area, calculated using Areimol from the Collaborative Computational Project (CCP4) software for Macromolecular X-Ray Crystallography

14B7 scFv variants with single alanines mutations were characterized and crystallized in order to study and fine-tune the energetics involved in the formation of a crystal lattice. Amino acids were selected for alanine mutagenesis with a wide range of altered free energy contribution to the interaction and contribute a large portion of their solvent accessible surface areas to the crystal contact surface area [196]. Amino acids previously reported as critical for loop conformation residues were eliminated as mutagenesis sites [197, 198]. The exception was light chain residue L:W33, which was included because more than 85% of the residue's exposed area was involved in the crystal contact.

Biophysical characterization of the variants showed fairly similar expression yields however a qualitative analysis showed a wide range of structural stabilities (Table 5.4). Size exclusion of antibodies demonstrated that most of the variants maintained their monomeric form in equilibrium with little cleavage (Fig. 5.2a and 5.2b), which was further confirmed with SDS-Page separation (Fig. 5.2c). CD measurements confirmed little structural perturbation in the variants with the exception of heavy chain variant W33A (Fig. 5.2d and 5.2e), corroborating previous literature which suggest this site is a structurally conserved residue [32]. Structural breakdown for each variant indicated traditional scFv formations of 8.5±3.9% alpha helices and 69.7±14.0% beta sheets (W33A not included in data set). Heavy chain variants R53A and D54A displayed unusual deviation in the calculated secondary structure composition, while maintaining similar characteristics to the WT in expression and thermal stability). The deviations are most likely the result of experimental error due to contamination from protein cleavage, unfiltered samples or noise at low wavelengths due presence of certain compounds like NaCl.

Table 5.4 Biophysical Characterization of 14B7 variants with alanine substitutions

Chain: Kabat#	Amino Acid (AA)	Δ ΔG in 14B7 (kcal/mol)[a]	Total Yield (mg)[b]	CD Anlysis (%α helices)[c]	CD Analysis (%β sheet)[c]	Thermal Stability (°C)[d]	Final Crystal Length (um)
WT	-	-	0.934	4.4	73.6	69.36 ± 0.36	200.0
L:28	D	-0.20	0.790	0.0	69.6		666.7 ± 291.7
L:30	R	2.66	0.897	9.5	87.0	58.27 ± 0.41	-
L:32	Y	2.17	0.617	7.0	75.7	54.93 ± 0.22	500.0
L:50	Y	2.06	1.115	9.2	79.3	56.37 ± 0.31	-
L:53	R	0.22	0.469	13.7	41.4	57.00 ± 0.82	125.0 ± 58.9
H:31	S	0.14	0.381	10.9	61.5	59.43 ± 0.058	375 ± 58.9
H:33	W	0.87	0.518	43.6	37.6	54.10 ± 0.44	-
H:54	D	0.65	0.627	10.2	54.3	59.00 ± 0.283	333.3
H:56	D	1.19	0.388	-	-	52.35 ± 0.64	-
H:97	L	1.16/0.66	0.655	11.8	74.4	58.53 ± 0.47	308.3 ± 82.5
H:98	L	2.18	0.572	11.7	88.3	57.60 ± 0.00	-
H:99	R	0.91	1.101	6.0	62.1	62.37 ± 0.23	358.3 ± 11.8

[a] Computational mutagenesis Alanine Scanning Robetta Software (average of all four antibodies involved in the two distinct contacts)
[b] From 250mL Culture of Terrific Broth and induced with IPTG for 4-5 hours
[c] Circular Dichroism Measurements were performed using Chen & Yang reference data set

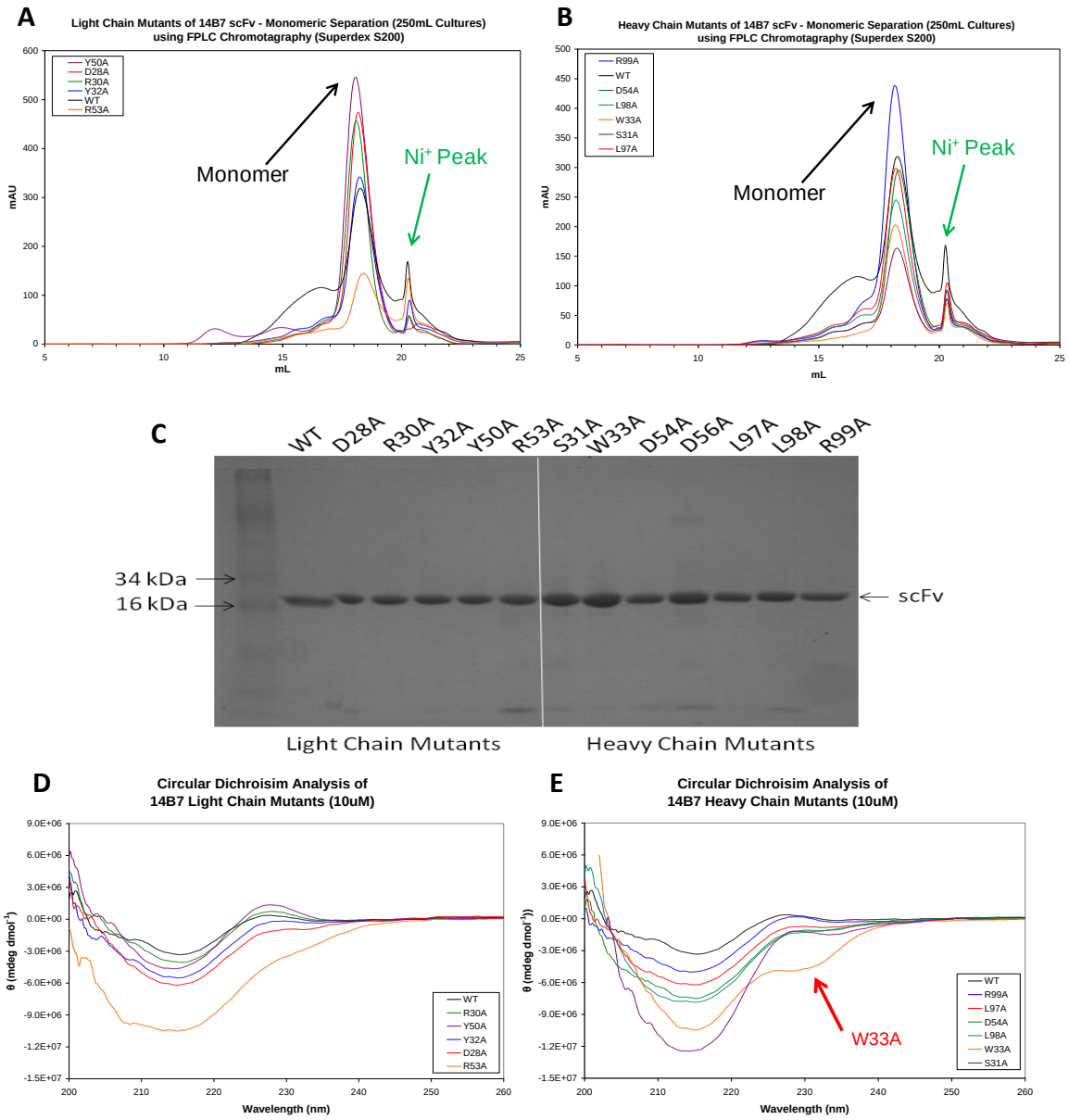


Figure 5.2 Biophysical Characterization of 14B7 variants

Top: Size exclusion purification (FPLC-SEC) of 14B7 and light chain variants (a) and the heavy chain variants (b) fractionated using superdex (S200) column.

Middle: (c) Separation of 14B7 and variants with 12% SDS-PAGE, stained with Pierce GelCode Blue. Single monomer band at expected peak at 28kDa for scFv, indication of a little cleavage of the antibodies.

Bottom: (d) Circular Dichroism (CD) measurements of 14B7 and light chain variants (e) CD Measurements of WT and the heavy chain variants.

5.4.2 The role of energetics in crystal contacts: crystallization of 14B7 variants

14B7 scFv single alanines mutants crystallized under the same conditions as the wildtype (20% PEG 4000, 0.1M Hepes (~ pH 7.5–8.5), 10% IPA) with varying crystal growth kinetics and produced different diffraction-quality crystal with lengths of 120 to 670um (Table 5.5 and Fig. 5.3). With the exception of light chain mutant D28A, crystal formations maintained the wildtype rod-like crystal shape, growing in one-dimension. Initial X-ray diffraction data indicated that most of the variants were high diffracting (~2.7 Å) with an initial index for the wildtype primitive tetragonal group (P4). Mutation at heavy chain D54A resulted in crystals that diffracted at low resolution (~15 Å), indicating a lack of long-range order within the crystal. 14B7 with a mutation at heavy chain D28A showed improved crystallization characteristics and will be discussed later in further depth.

Mutations at residues indicated by PISA to contribute either a hydrogen bond or an electrostatic interaction resulted in an impact on crystallization. The single alanines mutations at these particular residues resulted in either no crystal lattice formation (L: R30, Y50; H: W33, D56), modified crystal growth (L: Y32 – 3 month, H: S31 – 1 month) or poor diffraction quality (H: D54). The exception was H: R99, which indicates other forces may be involved with maintaining crystallization. The ability to impact 14B7 crystallization shows the potential of utilizing PISA to pinpoint critical amino acids that may be involved in generating stable intermolecular crystal contacts.

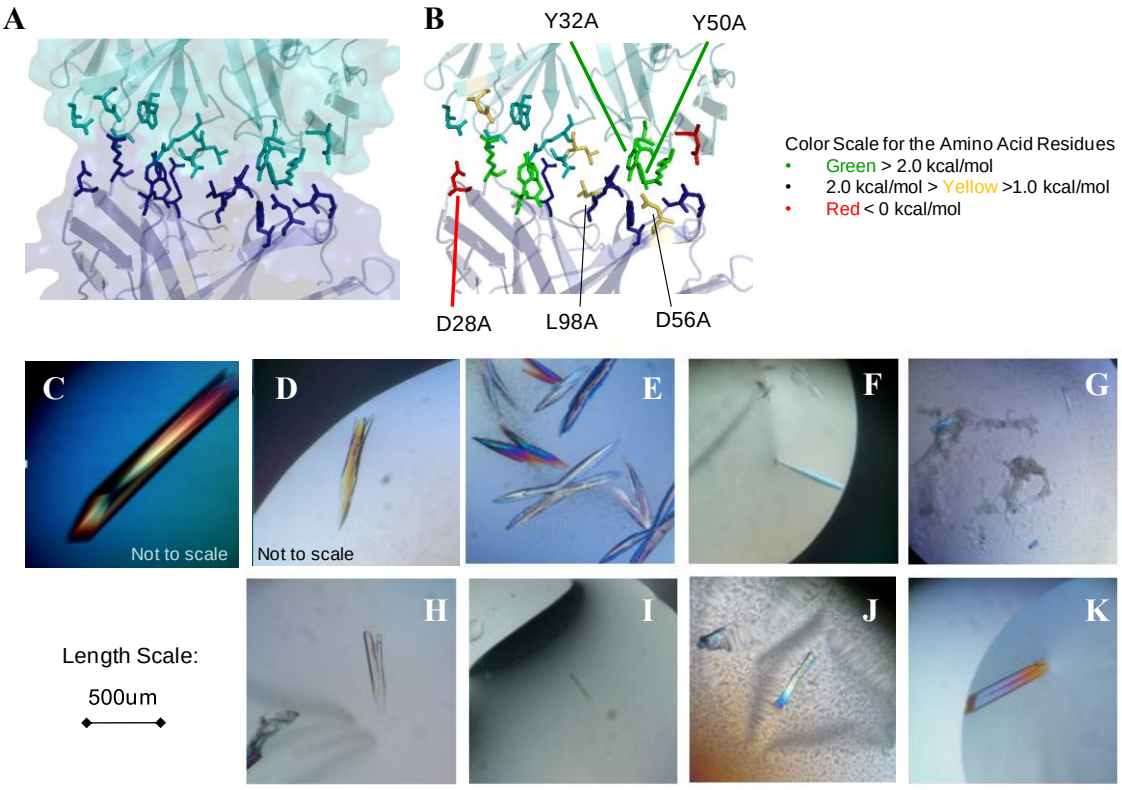


Figure 5.3 Crystallization of 14B7 mutants of the largest intermolecular contact

(a) Structural representation of the intermolecular contact;
(b) Location of all the amino acids subjected to experimental alanine scanning color-coded by their contribution of free energy; 14B7 and Variant Crystals at concentration of 15 mg/mL and a 1:1 concentration of mother liquor to protein within the sitting drop.
(c-k) Crystal Images of 14B7 variants with single alanines mutations at the largest contact
(c) WT-model image from previous work in the lab;
(d) WT; (e) L:D28A; (f) L:Y32A; (g) L:R53A; (h) H:S31A; (i) H:D54A; (j) H:L97A; and (k) H:R99A.

Table 5.5 Crystallization Characterization of 14B7 variants with alanine subsitutions

Chain: Kabat#	Amino Acid (AA)	Δ ΔG in 14B7 (kcal/mol)[a]	Final Crystal Length (um)	Crystal Appearance	P4 Unit Cell α=β=γ = 90° (a=b ,c)	Diffracting Crystal	Resolution (Å)
WT	-	-	200.0	24hr	80.18, 67.825	Y	2.0
L:28	D	-0.20	666.7 ± 291.7	2hr	84.31, 199.85	Y	3.2
L:30	R	2.66	-	-	-	-	-
L:32	Y	2.17	500.0	3 month	80.11, 68.21	Y	2.6
L:50	Y	2.06	-	-	-	-	-
L:53	R	0.22	125.0 ± 58.9	-	-	N	-
H:31	S	0.14	375 ± 58.9	1 month	-	N	-
H:33	W	0.87	-	-	-	-	-
H:54	D	0.65	333.3	24hr	N/A	N	13.0
H:56	D	1.19	-	-	-	-	-
H:97	L	1.16/0.66	308.3 ± 82.5	24hr	-	-	-
H:98	L	2.18	-	-	-	-	-
H:99	R	0.91	358.3 ± 11.8	24hr	74.79, 68.25	Y	3.7

Exploring the energetic contribution of these particular variants in more depth using Rosetta Alanine Scanning showed that mutations at residues with a free energy contribution ≥ 1.0 kcal/mol yielded no crystals following the 24 hours after initiation. In long-term crystallization studies, only variant Y32A managed to crystallize three months post-initiation. Based on the drastic change in growth kinetics indicates the Rosetta alanines threshold used for predicting critical amino acids in high affinity binding interactions is acceptable for determining residues involved in intermolecular contacts of crystallization [194]. Variants with residues that had a free energy contribution of < 1 kcal/mol showed varying crystallization kinetics. Residues L: R53, H: L97 and L98, which were specifically predicted only by the Rosetta program, conformed to the energetic criteria of Alanine scanning we have set for manipulating crystallization.

From the single alanine substitutions, it is apparent that modulating the energetics involved in intermolecular contact impacts crystallization. However, double mutations

were created to establish the extent to which this contact energetic is upheld and determine if a free energy contribution limit is needed to maintain the intermolecular interaction. All D28A double mutants (D54A, L97A, L98A, R99A) maintained a similar crystal growth rate to the single mutant D28A producing crystals within 2 hours, with the exception of variant: D28A + L98A. The negative free energy contribution of D28A appears to stabilize the intermolecular contact enough and allow for crystal initiation of the double variants. The loss of the free energy contributed by the heavy chain L98A, however appears to outweigh the stabilizing effect of the D28A mutation and results in no crystal formation.

Analysis of the intermolecular contact energetics established a total free energetic modulation of >1 kcal/mol at a protein-protein interaction impacts the crystal lattice formation for 14B7. The potential energy cost requirement for this particular intermolecular contact of 13.5kcal/mol must be reached to impact crystallization of 14B7 scFv. By fine tuning a key intermolecular contact, we have established a selection criterion for surface modification to control protein crystallization. PISA predicted residues that provide hydrogen bonds and electrostatic interactions potentially critical to lattice formation. Additionally, the free energy contribution analysis utilized for high binding affinity can be similarly applied to predict critical amino acids involved in low affinity intermolecular contacts. Residues critical to maintaining the interaction have a $\Delta\Delta G > 1$ kcal/mol and residues that stabilize the interaction have a free energy contribution < 0 kcal/mol. Utilizing the criteria for critical residues that impact intermolecular contacts, established by studying the 14B7 scFv crystallization kinetics, we can apply these to designing and improving the 3D5 based chaperones in developemont [Chapter 3].

5.4.3 Manipulation of surface epitopes on lattice formation

Due to its favorable characteristics, 3D5 scFv (PDB 1KTR) was used as a framework while engineering different anti-peptide chaperone variants in house. Modifications were introduced at varying locations on the HCDR regions and resulted in unique lattice formations for each engineered antibody under the same crystallization conditions. The surface epitope modifications resulted in the introduction of amino acids that contributed H-bonds or electrostatic interactions that generated additional driving contacts that affected the lattice space group the scFv crystallized into [Chapter 3]. Even with the overwhelming homology of these antibodies to the wildtype 3D5 (3D5/His_683 and 3D5/EE_48, >80%), these antibodies when crystallized under identical conditions resulted in their diverse lattice formations (3D5/EE_48: cubic space group, 3D5/His_683: hexagonal/tetragonal space group and 3D5: tetragonal space group). Examination of these scFvs will allow us to analyze the impact intermolecular energetics contributed by protein region modification can have on lattice formation.

3D5/His_683 scFv deviates from the wildtype only in the heavy CDR3 region (13 amino acids). In addition to crystallizing in the wildtype mother liquor solution, this scFv successfully crystallized in a number of other conditions that span a large range of pH 6.4 - 8.0 with varying buffers and different sizes of PEGs (Table 5.1). From these conditions, crystals were produced much more rapidly than the wildtype (within 4 days, compared to 2 weeks) in two different forms: teardrop and rod-like shapes. These crystallization characteristics suggest the potential of this scFv as a hyper-crystallizable antibody (Fig. 5.4). The 3D5/His_683 scFv has effectively been crystallized to lengths of 60μm with one crystal successfully diffracted to a resolution of 3.5Å. Preliminary diffraction of the crystal resulted in a densely packed space group ($P6_522$, 58% solvent) (Fig. 5.5c). Like the wildtype lattice formation, only one scFv molecule occurs in each

asymmetric unit and the preliminary analysis show most of the intermolecular contacts involve the framework. Further refining of this model is in progress to analyze the specific intermolecular contacts involved in these lattice contacts. Lattice comparison for each of the three 3D5 variants showed that all maintained a central cavity and the incorporation of the modified amino acids provided additional lattice contacts (Fig. 5.5).

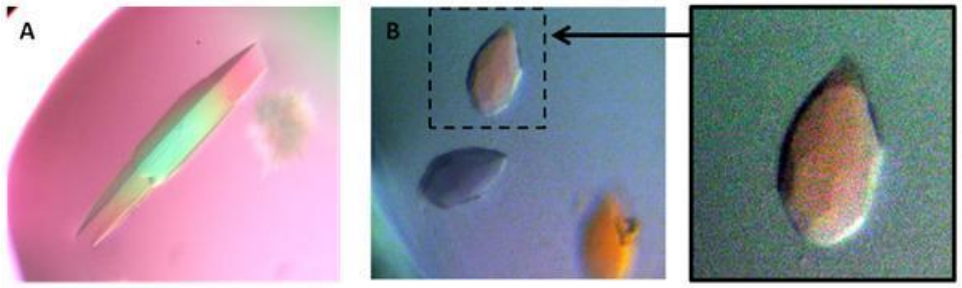


Figure 5.4 Images of 3D5/His_683 Crystals grown at different conditions

(a) rod shaped crystals
(b) tear-drop crystals with an enlarged image of one particular crystal

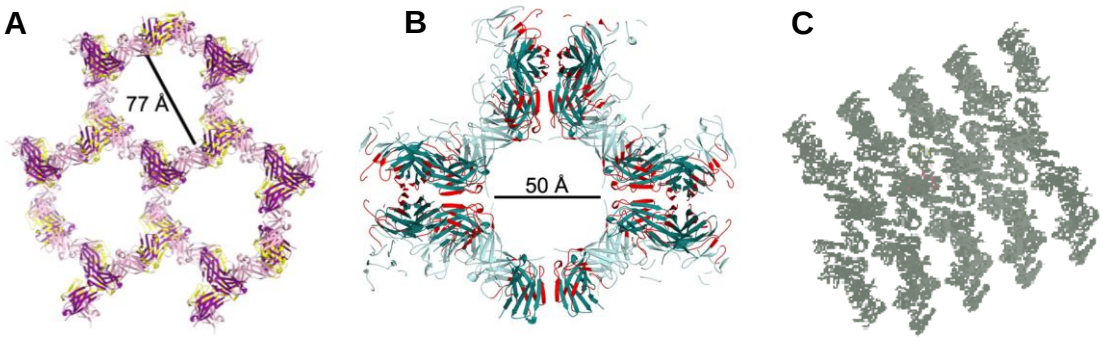


Figure 5.5 Comparison of 3D5 derived antibodies (3D5, 3D5/EE_48 and 3D5/His_683)

(a) 3D5 (purple) crystal lattices with size of central cavity indicated
(b) 3D5/EE_48 (blue) crystal lattices with size of central cavity indicated
(c) Preliminary crystal lattice for 3D5/His_683
For both structures, lighter hue indicates the V_L chain while the darker hue indicates V_H. CDRs for 3D5/EE_48 are depicted in red, 3D5 are yellow.

Further modification of the heavy CDRs in the anti-His_6 3D5 scFv resulted in crystallization of variant 3D5/EE_48 in the unique lattice space group F23 [Chapter 3] an extremely rare occurrence in protein crystals. A thorough analysis of the new lattice structure showed that the largest crystal contact was a retained intermolecular interaction from the 3D5 lattice. The next two largest crystal contacts in the 3D5/EE_48 lattice were novel interactions formed as a result of amino acids located at the modified HCDR regions. An energetic analysis using PISA revealed that the second largest contact (contact 2) was formed by hydrogen bonds contributed by heavy chain amino acids in CDR1: S30 and S32. The third largest contact (contact 3) involves salt-bridge interactions supplied by amino acids in the heavy chain CDR2 region: D55, D56, K57 and R58. The inclusion of these two new contacts appears to guide the F23 lattice formation. These contacts however are not beneficial to the design of our 3D5/EE_48

chaperone as they involve binding regions which should be accessible for peptide interaction.

Six variants were engineered to systematically modulate the energetic contribution of two novel intermolecular contacts using a PISA analysis of the energetic contribution of critical residues. These surface residues were modified to determine if lattice structure could be reverted to the wildtype's crystal formation. Each contact was removed individually as well as in combination to study the impact these had on lattice formation (Table 5.6). A variant designed for removal of contact 2 (3D5/EE_48.2a) resulted in crystallization while variants with contact 3 modified resulted in no crystal formations. The modification of contact 2 required two different variant versions because Ser32 was an inserted residue to extend the 3D5/EE_48 HCDR2 loop. As a result, two possible mutations were tested to remove this interaction i) alanine – a minimal substitution (3D5/EE_48.2a) or ii) threonine – similar characteristics with the -OH in a different location (3D5/EE_48.2). Effective modulation of contact 3 was accomplished by returning the critical residues (D55, K57 and R58) back to their original wildtype amino acids. Critical residue Asp56 was not included because of its potentially critical role in EE peptide tag affinity [Chapter 3]. Ser62 was included to restore a critical salt-bridge interaction that was present in the wildtype 3D5 lattice.

Biophysical characterization of 3D5/EE_48 variants showed proteins maintained a pure monomeric form in equilibrium solution, although modification at contact 3 negatively impacted protein expression and stability (Table 5.7, Fig. 5.6). Preliminary ELISA and Biacore analysis (not shown) showed reduced peptide affinity for variants with modified residues from the contact 3 region. While currently not a major concern, it must be taken into consideration as further work is continued development on this chaperone variant. Only variant 3D5/EE_48.2a has successfully been crystallized in the

same conditions as the 3D5/EE_48. Preliminary analysis indicates the crystal shape is similar. Diffraction data will provide more information regarding whether modification of this contact guided this lattice formation.

Table 5.6 Characterization of critical residues in 3D5/EE_48 scFv crystal lattice

Variant	Contact	Δ^iG (kcal/mol)[a]	Hydrogen bond Electrostatic Interaction[b]	SA in Crystal Contact (Å^2)[c]	% SA of AA in Crystal Contact[c]	% of Total Crystal Contact
S30	2	-0.14	H	54.45	91.2	10.5
S32	2	-0.17	H	68.72	88.7	13.2
D55	3	-0.18	S	10.73	13.5	2.2
K57	3	-0.60	S	47.81	69.6	10.0
R58	3	-0.68	S	76.07	49.3	15.9
S62	3	-0.07	-	6.17	10.2	1.3

[a]Solvation Energy Effect (PISA) – does not account for free energy contribution from the H-bonds or salt interactions
[b]Predicted using PISA, H- hydrogen bond contribution, S- Salt Bridge contribution
[c]SA – Surface Area calculated using PISA

Table 5.7 Biophysical Characterization of 3D5/EE_48 scFv variants with systematic contact manipulation

Variant	Total Yield (mg)[a]	Thermal Stability (^{o}C)[b]	Δ^iG removed (kcal/mol)[c]	S30T	S32T S32A	D55G	K57T	R58S	S62K
3D5/EE_48	4.5	50.25 ± 0.27	-	-	-	-	-	-	-
3D5/EE_48.2	3.1	51.67 ± 0.29	0.29	+	T	-	-	-	-
3D5/EE_48.2a	4.3	51.08 ± 0.21	0.29	+	A	-	-	-	-
3D5/EE_48.3	0.5	58.08 ± 0.20	1.84	-	-	+	+	+	+
3D5/EE_48.23	3.5	59.00 ± 0.55	2.13	+	T	+	+	+	+
3D5/EE_48.23a	0.9	58.33 ± 0.26	2.13	+	A	+	+	+	+

[a]Based from 1L culture
[b]Using RT-PCR
[c]Solvation Energy Effect (PISA) – does not account for free energy contribution from the H-bonds or salt interactions

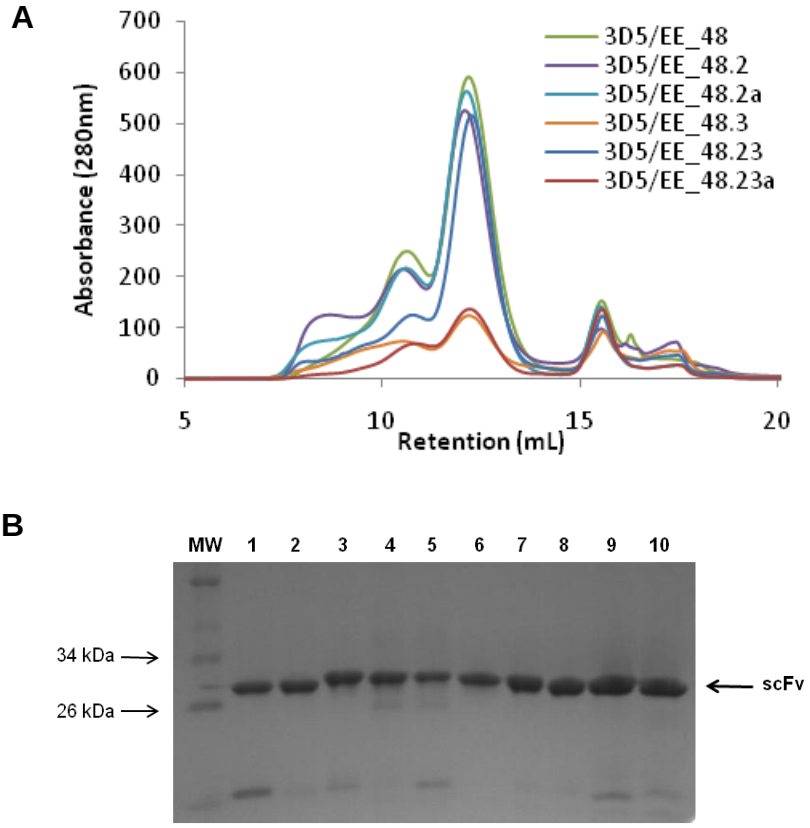


Figure 5.6 Biophysical characterization of 3D5/EE_48 variants

(a) Purification by Size Exclusion Chromatography on S75 Column. Molecular weight standards on S75 elute at 7.4 ml (void volume; blue dextran 2000); 9.6 ml (75 kDa conalbumin); 10.9 ml (43 kDa ovalbumin); 12.7 ml (29 kDa carbonic anhydrase); 15.4 ml (residual peak from Ni^{2+} affinity column).
(b) Size and purity of scFv by SDS-Page. Lanes are 1: 3D5, 2: 3D5/His_683, 3: 3D5/His_67, 4: 3D5/EE_48, 5: 3D5/EE_51, 6: 3D5/EE_48.2, 7: 3D5/EE_48.2a, 8: 3D5/EE_48.3, 9: 3D5/EE_48.23, 10: 3D5/EE_48.23a

The surface epitope modification of 3D5 to create or eliminate multiple critical intermolecular contacts allows better understanding of its impact on lattice formations. From our global study of 3D5 framework antibodies, the overall propensity of a protein for crystallization seems to be driven by key intermolecular contacts. The introduction of these amino acids appears to have provided an unintentional new lattice contact. The

systematic modification in the HCDR3 suggests a new interaction was created that guides the hyper-crystallizability of 3D5/His_683. Further modification of all the heavy chains contributed to multiple protein-protein interactions and resulted in the formation of a completely unique space group for 3D5/EE_48, indicating the role multiple intermolecular contacts have in affecting lattice construction. It appears that in this family of antibodies the modulation of one intermolecular contact can impact the lattice structure of the crystal, indicating that more than one protein-protein interaction is critical for initiating crystallization. Examination of particular driving contacts in 3D5/EE_48, allow us to explore the key intermolecular contacts involved in the crystallization propensity of this particular antibody. Other 3D5 framework antibodies that carried similar homology (>80% homology, in PDB) can be analyzed for their crystallization characteristics for potential contacts. By characterizing these contacts in detail, we can modulate specific low affinity interactions in 3D5/EE_48 without impacting critical high affinity peptide binding regions. This provides the potential to create a hyper-crystallizable antibody with maintained functionality for potential use as a crystallization chaperone.

5.4.4 Application for engineering enhanced chaperones for protein crystallization

Current protein engineering techniques are used to directly modify and improve the particular target protein crystallization propensity. Numerous proteins have been successfully enhanced for protein crystallization by improving different biophysical characteristics. The two most popular methods target improving solubility or reducing the surface entropy of the protein through targeted point mutations. Rational enhancement of protein solubility has successfully improved crystallization of other

proteins like S1 hydrofolate reductase [202], Leptin [203], mitogen-activated protein (MAP) Kinase [204]. Successful amino acids recommended for improved solubility include mutations to Arg and Ser, as well as mutating Asn and Gln to their acids (Asp and Glu) [192].

Effective reduction of the surface entropy of the target protein (SER) has helped to crystallize several novel proteins that had been unsuccessfully crystallized in their wildtype form [66, 67, 205]. Surface topology of proteins contributed by flexible amino acids such as Lys, Glu and Gln can interfere with intermolecular interactions that allow molecules to associate together and are necessary for crystallization [67, 177, 205]. Systematic removal of these high entropy residues with smaller residues like alanines, was most effective when multiple amino acids clusters are simultaneously mutated [68]. This generates a more symmetric unit for crystal packing. This success of this strategy has resulted in the development of servers that currently assist in identifying the critical clusters for mutation [70]. Both of these processes are effective at improving crystallization. However, our understanding of the energetics involved at the intermolecular contact can provide additional guidance for facilitating the creation of improved intermolecular contacts. This energetic criterion in conjunction with the other techniques can provide a more comprehensive targeted modification of the surface topology to develop improved energetically stable lattice interactions.

Like SER, modification of solvent accessible residues utilizing the energetic criteria requires structural information, which is especially challenging for proteins not successfully crystallized. This limits the number of proteins that these techniques can assist in crystallization. For our platform peptide specific chaperone technology this is not the case as the antibody for modification is characterized with atomic resolution maps. The use of a peptide specific chaperone drastically reduces this number of

chaperones need to crystallize the wide range of proteins. As a result, it directs the focus on these particular chaperones to engineer, study and improve the low affinity interactions that guide crystalline growth. These chaperones serve to assist the target protein in crystallization by mediating the lattice forming contacts for the complex. Our detailed analysis of the interaction energetics of the 3D5 based antibodies will guide us as we fine-tune their intermolecular contacts for improved crystal propensity. Our energetic selection criteria can be utilized to select critical residues in lattice formation on the 3D5, 3D5/His_683 and 3D5_48 scFvs for improved lattice formation. The design of this hyper-crystallizable chaperone will provide a generalizable approach to enhancing crystallization of difficult proteins.

5.5 CONCLUSION

Engineering a hypercrystallizable peptide specific chaperone is the critical design component for a platform chaperone-assisted crystallization technology. The ability to improve the chaperone protein's propensity to crystallize is of utmost importance. In order to modulate the chaperone for hyper-crystallizable characteristics a better understanding of the role of lattice contacts in crystallization is needed. The two different murine model family antibodies were used to examine global effects of crystal contacts on lattice formations and the small scale effects of individual residues on a specific protein-protein interaction. The protein surface modifications indicate the role intermolecular contacts play in defining crystallization.

Programs like Rosetta alanine scanning that calculate energetics for high affinity interactions can similarly be applied to select residues in low affinity interactions of crystal contacts to modulate crystal lattice formation. Variations in these intermolecular

contacts impact the lattice formation that is generated. Further work is required to fully understand the impact of surface modification on intermolecular interactions, crystal lattice mechanisms and assembly kinetics. Utilization of the information will assist in improving the crystallization propensity of our 3D5/EE_48 chaperone by engineering thermodynamically viable specific interactions.

CHAPTER SIX: Conclusions and Recommendations for Future Direction

6.1 CONCLUSION

The dissertation presented here is composed of the necessary steps towards achieving a general platform for crystallizing target proteins. High quality protein crystals have previously been obtained using protein-specific antibodies as a co-crystallization agent in complex with the protein of interest. The identification and optimization of each protein specific chaperone becomes expensive and time-consuming. Even after obtaining a potential chaperone, optimal co-crystallization conditions must be determined for the complexed chaperone and target protiein. Our platform seeks to generalize this system with the development of a hyper-crystallizable antibody fragment with binding specificity for a short peptide sequence, which can easily be inserted into the target protein of interest at functionally silent regions without destroying the structural stability. In order to study the feasibility of this platform we first must: (1) engineer a peptide-binding scFv [Chapter 2], (2) characterize the potential scFv chaperone [Chapter 3] and (3) optimize the peptide ligand location to enhance complexation [Chapter 4] and (4) study the scFv's crystallizability for potential co-crystallization [Chapter 5].

We developed a novel and efficient method to generate scFv antibody libraries for improved solubility, expression and crystallization (MegAnneal). With this method, we bypass a major hurdle in recombinant DNA by circumventing the restricting ligation step to produce libraries rapidly and with large population numbers [Chapter 2]. Utilizing this technique and the anti-His 3D5 murine framework, the intention was to retain framework crystal contacts while modifying the peptide specificity. From our different 3D5-based

phage display libraries, we demonstrated the ability to select a number of antibody fragments with specificity for the His_6 tag and one scFv with successful conversion to EYMPME (EE) tag specificity, while enhancing biophysical characteristics (improved stability, solubility and monomeric expression levels). The resulting anti-EE antibody, 3D5/EE_48, possessed higher affinity (K_D: 26nM) than the variants with His_6 tag affinity and readily formed complexes with proteins tagged with peptides at various locations (terminally and internally) [Chapter 3]. Crystallization of 3D5/EE_48 produced an unexpected crystal lattice, resulting in the creation of two new contacts while retaining a previous contact from the original antibodies crystal lattice. Although the lattice is in a different space group (F23 vs. 3D5's $P3_221$) it still retains a central cavity (~50Å) that carries potential for crystal scaffolds.

Upon establishing 3D5/EE_48 as a viable co-crystallization chaperone candidate, the focus shifted to optimizing the location of the peptide: N-terminal, C-terminal and internal [Chapter 4]. To effectively create a co-crystallization scaffold, the peptide must be optimized for tight chaperone binding resulting in the formation of rigid complexes with minimal structural perturbation to the target protein. Terminal tags linked to a protective tail provide a good affinity (K_D: 389nM) and are a preferred location for peptide insertion since no structural information is necessary. Two internal locations were selected to study the impacts EE-internalization has on structural stability, the first in natural loops and the second in artificial linkers. Maltose binding proteins with direct EE tag replacement in a natural loop produced an accessible peptide for the chaperone to form a rigid complex. Insertion of a single peptide tag into the interdomain regions of the scFv required a lead sequence (~4 amino acids) in order to achieve the correct structural stability and flexibility in conformation for chaperone accessibility.

Crystallization results from engineered 3D5 derived scFvs indicate the need to better understand the forces that guide crystal lattice formation [Chapter 5]. Surface modifications of 3D5 resulted in novel and unique lattice formations (3D5: tetragonal, 3D5/EE_48: cubic, 3D5/His_683), with 3D5/EE_683 scFv variant displaying hyper-crystallizable characteristics. The new crystal contacts appear to involve modified amino acids in the HCDR regions and are contributing additional protein-protein mediated contacts, which is impacting protein crystallization. Energetic modulation of key intermolecular contacts impacted crystal growth. Systematic modulations allowed us to establish the role energetic contribution of intermolecular contacts play in crystal lattice formations. The energetic analysis of high affinity intermolecular contacts with programs like Rosetta Alanine Scanning, can be similarly applied to low affinity intermolecular crystal contacts. Understanding the role of intermolecular contacts in crystallization will assist in guiding the design our hyper-crystallizable chaperone antibodies and improving the diffraction quality of its crystals.

6.2 Recommendations for Future Direction

The successful characterization of the 3D5/EE_48 scFv with tight binding to the EE epitope at various regions demonstrates the potential for this antibody as a crystal chaperone. The in depth analysis of its biophysical properties has been explored, however, the exact interaction between the ligand and the antibody has not been determined. The generated model of the interaction, using computation docking (ClusPro), provides a possible prediction however, high diffraction quality co-crystals of 3D5/EE_48 scFv with EE-peptide are needed for structural determination. Additional co-crystallization studies with peptide-tagged proteins are also needed to effectively study

the resulting complex between the chaperone and the POI for rigidity and high affinity. Currently co-crystallization trials with 3D5/EE_48 scFv and select EE-tagged proteins are underway (R30A-EE_2, MBP-EE, MBP-EE-his, SPP-EE), however other proteins must be added. Other proteins to include are the MBP variants with internalized EE tag at the interloop region (amino acid 171 – 176) and the TCR with single EE tags at the interdomain linker. Successful assisted crystallization of these EE-tagged proteins will provide additional support that this scFv carries the potential to serve as a co-crystallization chaperone.

The development of 3D5/EE_48 with high affinity for EE-tagged proteins opens the possibility of engineering other peptide specific antibodies for use as crystallization chaperones. This creates a limitless number of future projects, which involves the creation of multiple peptide specific hyper-crystallizable chaperones derived from similar frameworks. By switching out the 3D5 CDRs for specificity to other short peptide epitopes, will lead to the development of a potentially generalizable toolbox of co-crystallization chaperones. 3D5/His_683's hypercrystallizable characteristics suggests it's potential as a chaperone and work should be done to improve its His_6 affinity. Efforts have also focused on the Flag tag (DYKDDDK) due to the presence of tyrosines and charged residues that favor hydrophobic and electrostatic interactions, respectively. The CDRs that bind this tag also display 10-fold higher affinity to another tag (MDYKAFDNL) [180]. It is therefore recommended to graft these CDRs onto the 3D5 framework, generating a new chaperone with the ability to bind these two different short peptides. Ambiguous HCDR3 residues in the PDB, resulted in unsuccessful preliminary attempts at producing the graft with high affinity and high yields (courtesy of Kevin Entzminger). It is recommended that engineering this anti-Flag antibody use site directed and random mutagenesis simultaneously. This can be accomplished with the MegAnneal

technique by amplifying 3' directional megaprimers under error-prone conditions utilizing a primer that contains degenerate codons at the H:CDR3 region. This methodology will rapidly select for both a soluble, high expressing and tight binding chaperone with 10X flag affinity using a single library. With the successful selection of 3D5/EE_48 using this method, there is high expectation that an improved 3D5/Flag variant will be found.

The chaperone assists in protein crystallization by contributing additional polar surface for lattice-forming crystal contacts, locking down flexible regions for increased overall uniformity and stabilizing the target protein for a homogenous solution. Our studies primarily focused on the scFv chaperone antibody format due to its small uniform shape and its ability to produce a more symmetrical and compact complex. It is questionable whether a single scFv (28kDa) will effectively provide enough surface for a target protein (50 – 100kDa). Two ways that we propose to study and address this problem involve: i) using a larger antibody chaperone format or ii) inserting multiple peptides at various regions in the POI. If the hyper-crystallizable trait of our 3D5/EE_48 scFv can be transferred into this Fab antibody format (50kDa) with similar lattice contact formation, this candidate can provide additional polar surface contributions making this a more advantageous potential chaperone. The Fab has successfully been cloned into the pFabs plasmid provided by the Georgiou lab with characterization and co-crystallization of these antibodies in progress, similar to those previously mentioned for the scFv version of the chaperone (Appendix A).

Another way to provide additional polar surface, is to surround the target protein with multiple scFv chaperones. We hypothesize that the use of multiple peptides and multiple chaperones would be advantageous for creating and controlling an intended scaffold system. This can be accomplished utilizing one of two options: i) the same

peptide in multiple regions or ii) multiple peptides in different regions (Figure 6.1). Our engineered MBPs displaying internalized EE tags [Chapter 4] have already been designed with the second suggestion in mind. Located at the c-terminal regions of these proteins, the ten times higher affinity Flag Tag (MDYKAFDNL) has been inserted prior to the His_6 tag. The use of two distinct chaperones will potentially allow for further control in the formation of the lattice structure by utilizing their distinct driving contacts. Once a hyper-crystallizable chaperone with affinity for the Flag tag is found we can proceed with the complex isolation studies of the correct chaperones and target protein in addition to co-crystallization studies.

The generation of additional 3D5 based chaperone antibodies will also help us to further explore the driving forces involved in protein crystallization for this particular family. With the recent successful crystallization of 3D5/His_683, a detailed analysis of its crystal contacts in the different crystal shapes will provide understanding of the driving force behind this protein's propensity for crystallization. Comparing the intermolecular contacts of the other scFvs in the family (i.e. wildtype 3D5 and 3D5/EE_48) will provide a more detailed analysis of the 3D5 family and further expand our knowledge of key intermolecular contacts for our chaperone. Our current understanding of crystallization has focused primarily on only two families of antibodies: 3D5 and 14B7, and provided some observations of how their intermolecular interactions impact lattice formation. The recommended next step is to expand our survey to analyze other families of single chain fragments (>80% homology) in the PDB database. In doing so, we would like to analyze the sequence properties (i.e. mean hydrophobicity, net electrostatic charge, side chain entropy, etc.) to determine the role surface epitopes play in generating intermolecular contacts. This data will provide a more thorough and detailed understanding for the sensitivity of scFv intermolecular interactions on a

protein's propensity to crystallize. The result: a more systematic guideline for designing and engineering our chaperones for hyper-crystallizability.

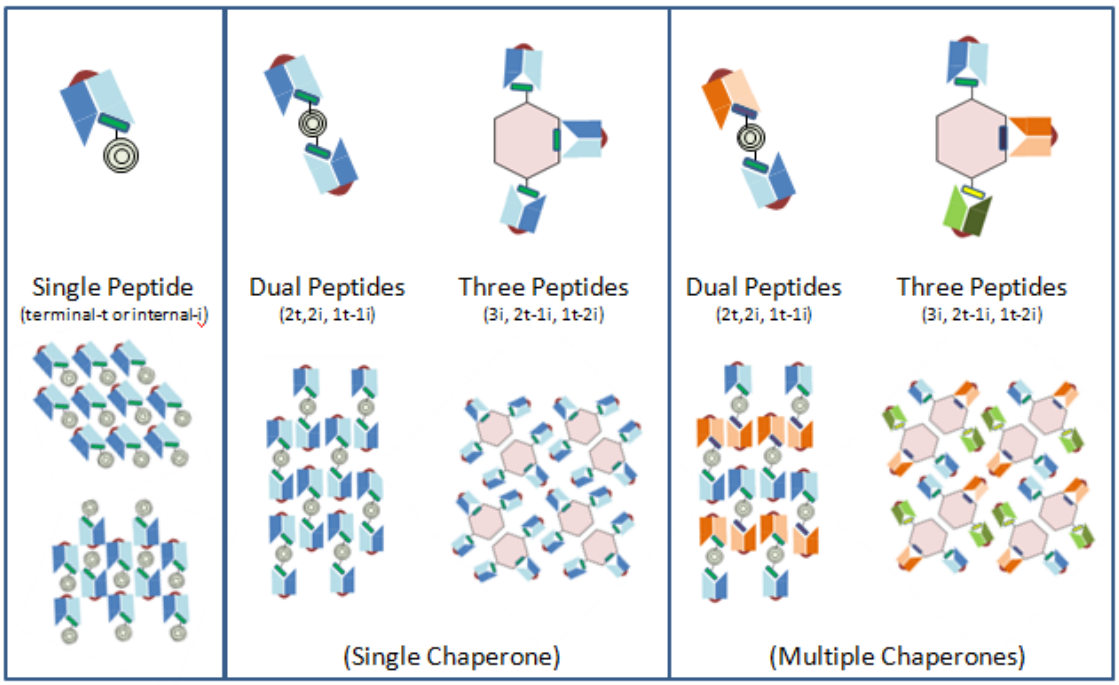


Figure 6.1 Depiction of the potential application for the toolbox of multiple scFv chaperones specific for different peptides.

Appendices

APPENDIX A: OTHER PROJECTS IN PROGRESS

A.1. Producing Fab version of 3D5/EE_48

The V_L and V_H were subcloned into the pFabs plasmid provided by Dr. George Georgiou (refer to Appendix B). A His_6 tag is subcloned into the light chain portion of the Fab while a Flag tag is located at the end of the heavy chain portion for purification purposes. The protein was expressed with BL21 E. Coli at 37^{o}C for 14 hours before incubating at 25^{o}C with 1mM induction of IPTG for 5 hours. Periplasm isolation was performed as mentioned in Chapter 3 with Ni^{2+} Affinity Chromatography followed by Size Exclusion with an S75 Column (Figure A.1a). The yield produced from a 1L cell culture was 1.77mg comparative to the 2.1mg generated from the scFv format of this antibody.

Biophysical characterization using 12% SDS-PAGE of the Fab protein with non-reducing buffer indicated that the solution maintains a largely monomeric form however there are some byproducts (Figure A.1b). Our purification methods involved using only metal affinity, which targets the light chain regions, which may explain the shorter bands found as the rentention volume increases. The shorter bands suggest that only the V_L region was selected indicating that a second purification step (such as anion exchange chromatography) is needed that involves the Flag tag for pure Fab separation. ELISA utilizing protein from the 10.5-11mL (A12) Retention fraction suggest that affinity for the EE tag was retained and is comparable to the scFv format.

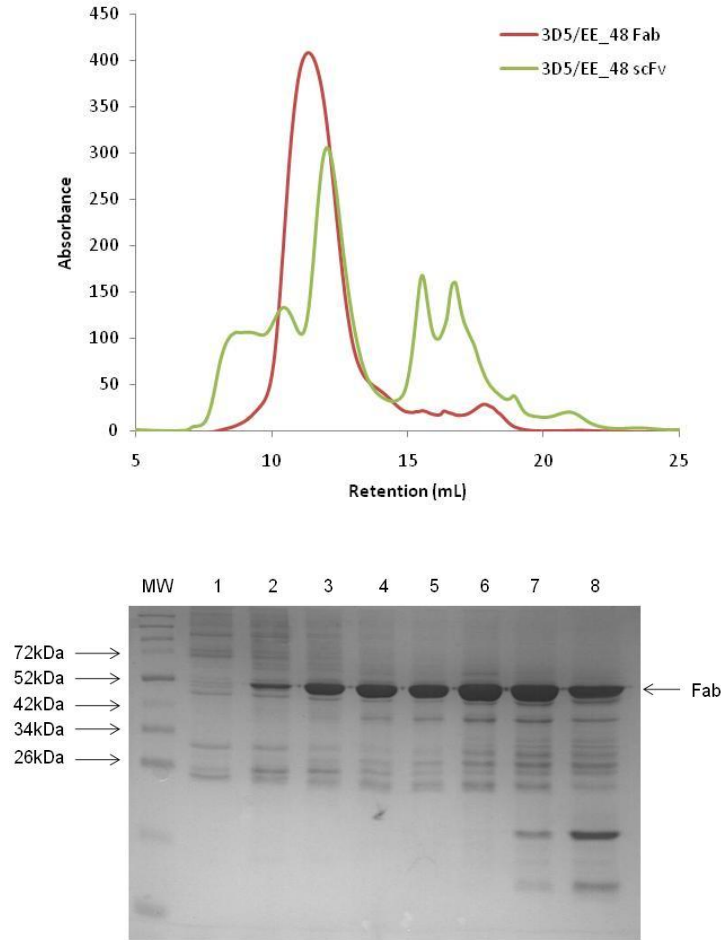


Figure A.1 Biophysical Characterization of 3D5/EE_48 Fab

(a) Size Exclusion Chromatography Purification comparison of the 3D5/EE_48 Fab format and scFv format

(b) Size separation using 12% SDS-Page with non-reducing buffer. Samples were collected at different flow-throughs off the SEC Lane 1: 9.5-10mL (A10), Lane 2: 10-10.5mL (A11), Lane 3: 10.5-11mL (A12), Lane 4: 11-11.5mL (B12), Lane 5: 11.5-12mL (B11), Lane 6: 12-12.5mL (B10), Lane 7: 12.5-13mL (B9), Lane 8: 13-13.5 (B8)

A.2 Engineering humanized 1B7 with improved solubility and expression yields

Below is a slightly more detailed discussion library generation with MegAnneal of humanized 1B7 which has high affinity for the pertussis toxin. The library (10^6 transformants) displayed 1B7 scFvs with on average 3 mutations per gene. Preliminary screening using phage-display produced one particular gene of interest, which is currently being subcloned into pak400 and pMopac54 for protein expression and characterization.

As mentioned, 1B7 had some processivity issues and primer annealing issues, requiring some modification to the 3' megaprimer generation process. Both mutazyme and the 5' pakpel primer were ineffective in generating the correct PCR product size (750bp), therefore production of the 3' megaprimer involved PCR with a biased nucleotides, Taq polymerase and only the 3'Huck primer. A strong PCR product at 750bp was excised and purified for use as the 3' megaprimer. A preliminary ELISA colony screen showed some potential candidates even with the relatively low growth of protein presenting phage (Figure A.3). Note: ELISAs were read before HCl neutralization (i.e. low absorbance signal). Panning over pertussis toxin (PTX) will be the next step for successful selection of a high expressing humanized 1B7 scFv.

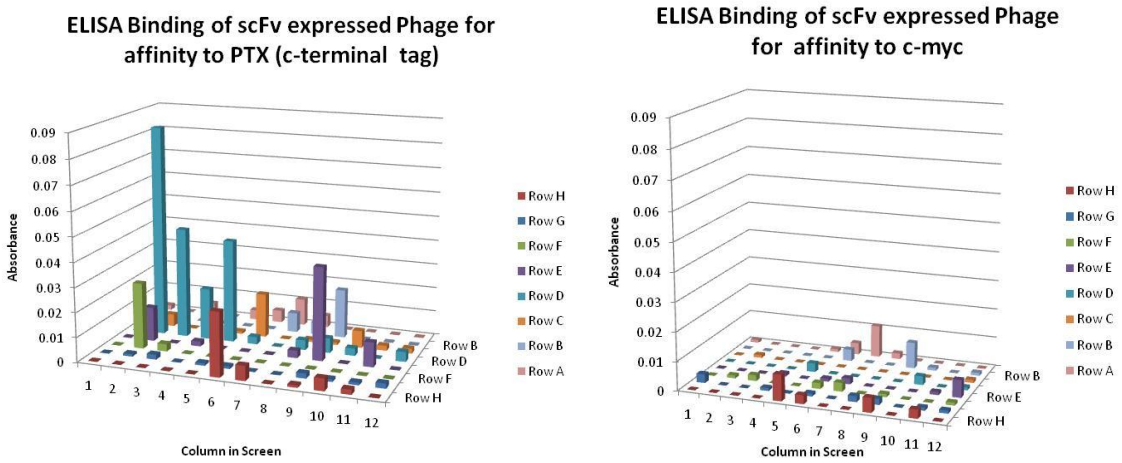
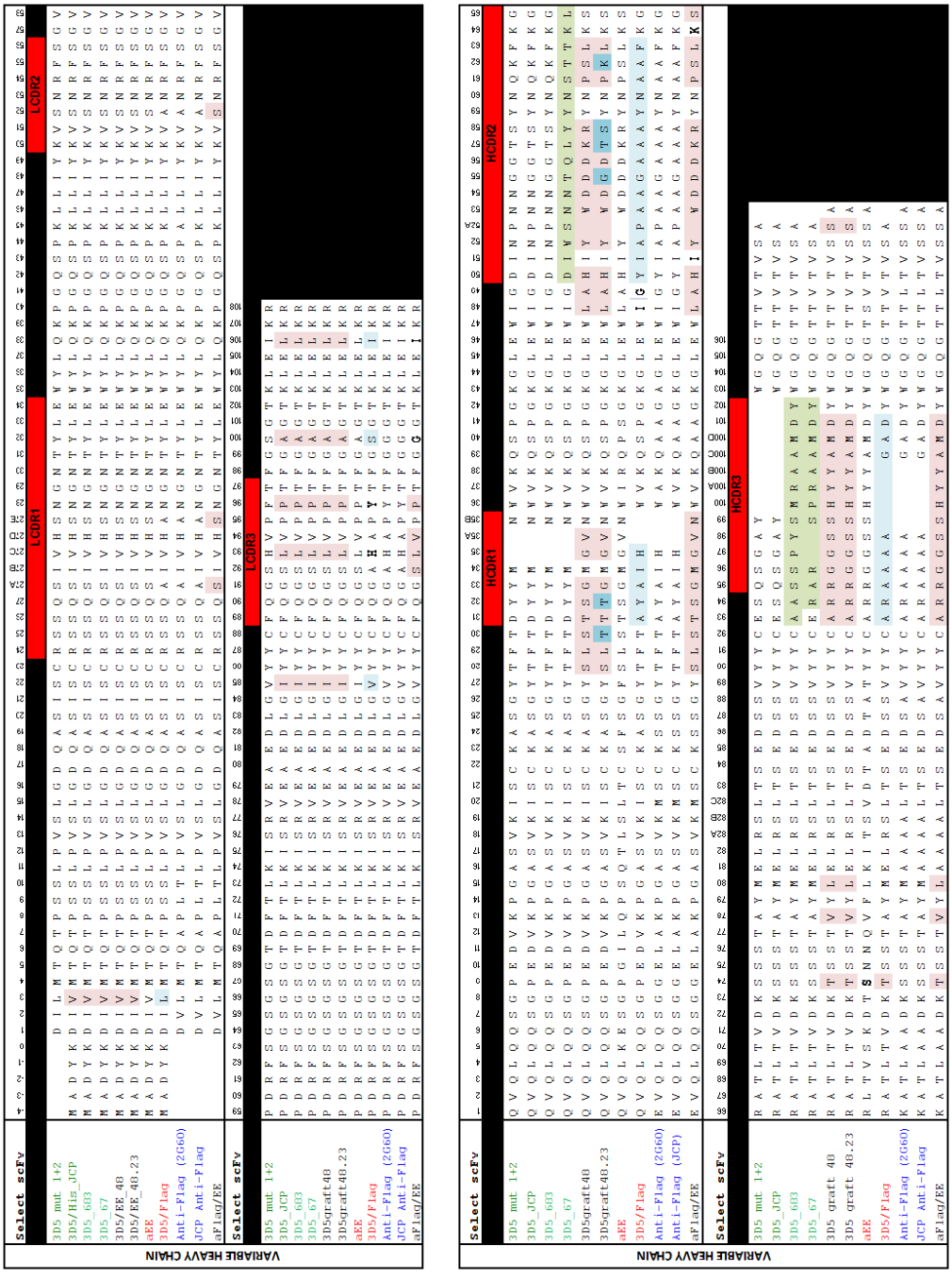


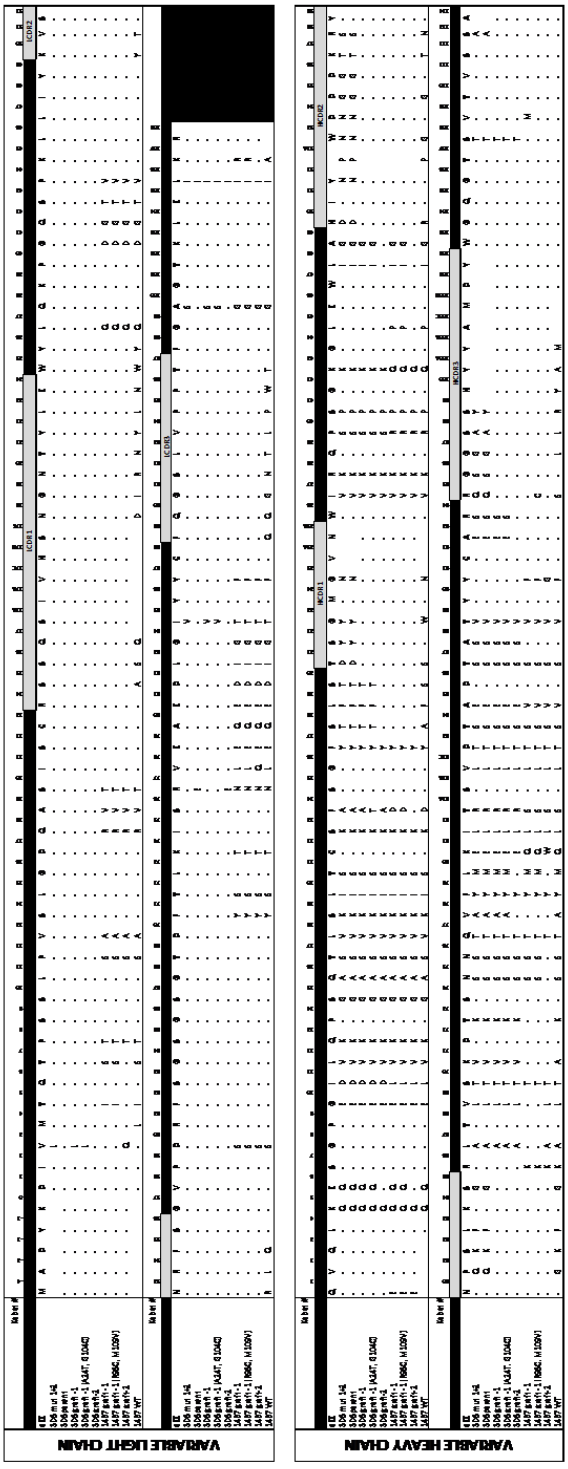
Figure A.2 ELISA Colony Screen of select hu1B7 derived from the library generated.

APPENDIX B: ADDITIONAL FIGURES FOR WORK PRESENTED

B.1 Amino Acid Sequence for Engineered 3D5 antibodies

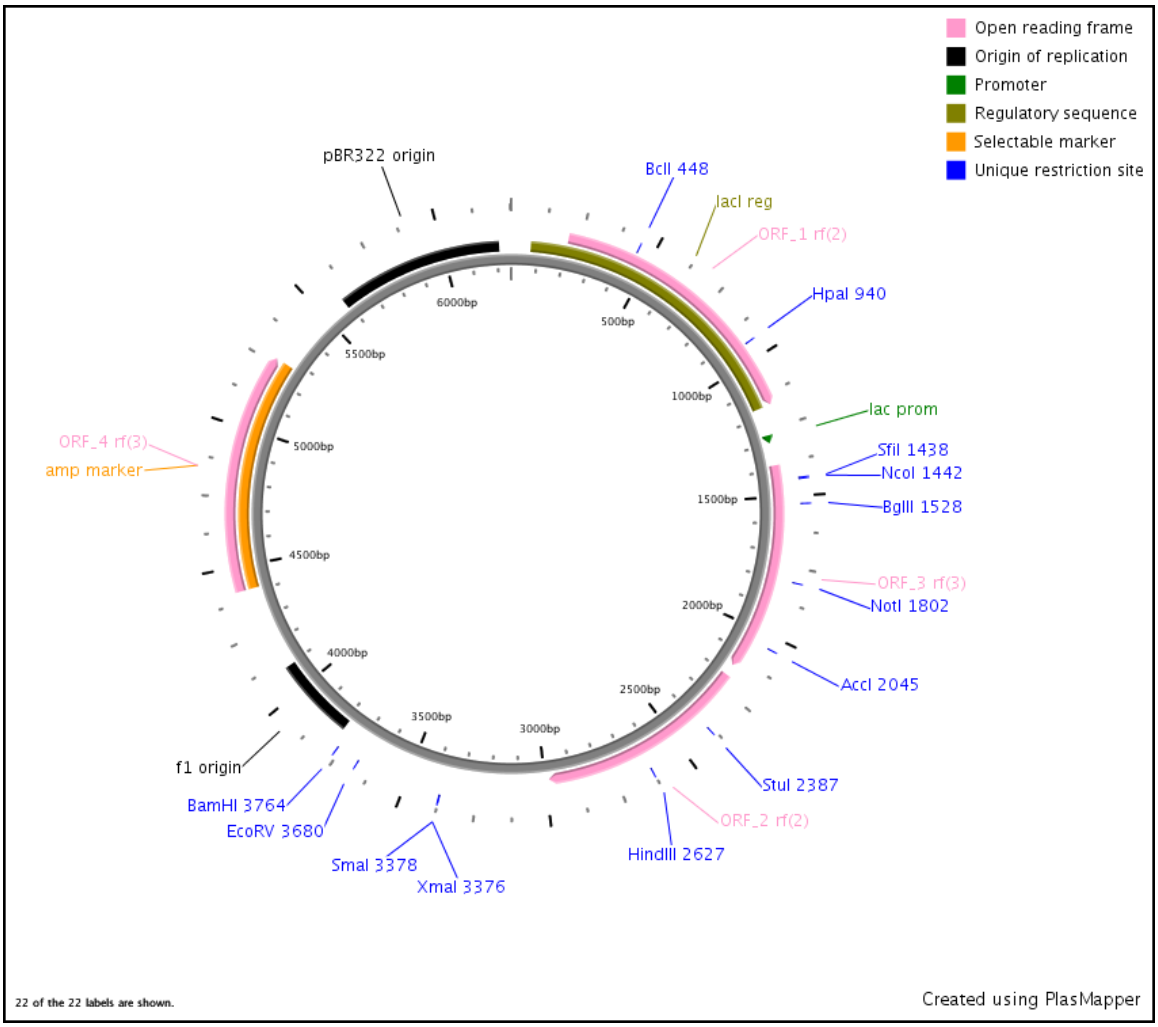


B.2 Amino Acid Sequence Alignment for Engineered Grafts



B.3 Vectors Maps Used for Periplasm Expression

B.3.1 pFabs Plasmid (Courtesy of G. Georgiou)



Typical lac promoter and expresses 1-5mg/L culture

VL – inserted between Nco1 and Not1 sites (1440 – 2124 – His_8)

VH – inserted between Nhe1 and Hind111 sites [Flag tag at Heavy Chain end] (2268 – 2980 – Flag)

NcoI	(C'CATG,G)	37°C with NEB 1/2/3/4	1442
NotI	(GC'GGCC,GC)	37°C with NEB 3	1802
NheI	(G'CTAG,C)	37°C with NEB 1/2/4	2262
HindIII	(A'AGCT,T)	37°C with NEB 2	2627

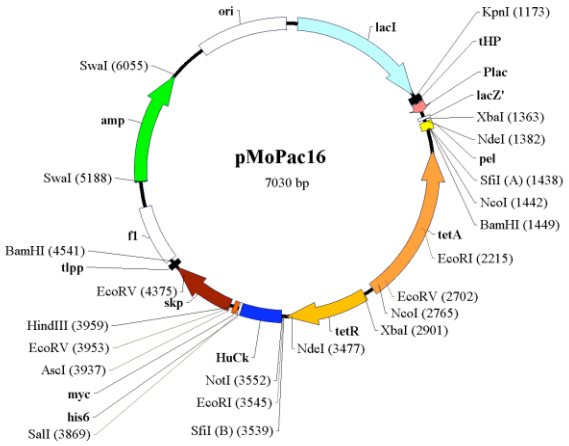
Reading Primers: 5’ pFabs_1309 (VLfor): GTG AGC GGA TAA CAA TTT CAC ACA GG

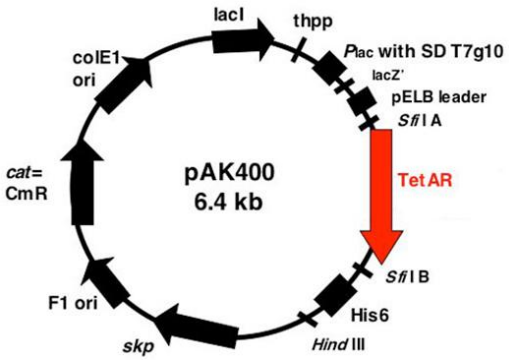
LENGTH: 26
GC CONTENT: 46.2 % MELT TEMP: 58.2 ºC
MW: 8028.3 g/mole EXT COEF: 263600 L/(mole·cm)
nmole/OD260: 3.79 µg/OD260: 30.46

5’ pFabs_2184 (VHfor): GAG AAG GAG ATA TAC ATA TGA AGT CGC TAC TCC C

LENGTH: 34
GC CONTENT: 44.1 % MELT TEMP: 59.8 ºC
MW: 10483.9 g/mole EXT COEF: 346800 L/(mole·cm)
nmole/OD260: 2.88 µg/OD260: 30.23

B.3.2 Other Common Vectors Used in Lab:



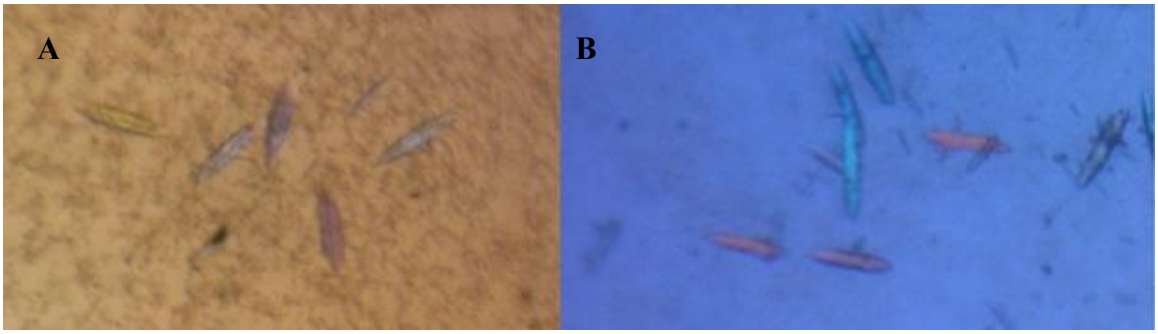


B.4 Crystallization Conditions for the Antibodies Used in Crystallization Trials in this Work

Antibody	Crystallization Conditions (Mother Liquor)	Protein Conc. (Protein:ML)	Crystal Appearance	Reference
14B7 (WT) scFv	0.1 Hepes (pH 7.5) 20% PEG 4000 10% IPA	~14mg/mL (1:1 or 1:2)	24 hours	Maynard
14B7 (M18) scFv	0.1M HEPES (pH 7.5) 20% PEG 4000 10% IPA	~14mg/mL (1:1)	6 months	Harvey
3D5 (mut1+mut2) scFv	0.1M MES (pH 6.4) 0.2M MgAc 20% PEG 8000 0.02% Sodium Azide (NaN_3)	~3.8mg/mL (5:1)	2 weeks	Kauffman
Anti-EE scFv and scAb	0.1M Tris (pH 8.5) 1.0M Lithium Sulfate 0.01M Nickel (II) Chloride Hexahydrate	~15 mg/mL (1:1)	4 months	In House
3D5/EE_48	0.1M Mes (pH 6.4) 0.1 MgAc 20-24% PEG 8000 – 4°C	~3.8 mg/mL (1:1 or 1:2)	4 weeks	In House
3D5/His_683	**Still in Progress**			

B.5. Crystallization of αEE scAb monomers (Images)

(a) and αEE scFv monomers (b) crystallized with conditions: 0.1M Tris (pH 8.5), 1.0M Lithium Sulfate, 0.01M Nickel (II) Chloride Hexahydrate.



B.6 Primers Utilized in 3D5 Library

(3D5 parent actual sequence in pM024 and pak100 frozen stock!!)

B.6.1 Stop Codons into Template ss-DNA

```
  P  G  K  G  L  E  W  I  G  D  I  N  P  N  N  G  G  T  S  Y  N  Q  K  F  K  G
                                                          TAA   TAA      TAG
GTCCAGGAAAGGGTCTGGAGTGGATCGGAGACATTAACCCGAATAATGGCGGCACGAGCTATAACCAAAAGTTCAAGGGC
CAGGTCCTTTCCCAGACCTCACCTAGCCTCTGTAATTGGGCTTATTACCGCCGTGCTCGATATTGGTTTTCAAGTTCCCG
                                        GCTTATTACCGCCGTGCTCGATTTTGATTTTCAAGATCCCG

R  A  T  L  T  V  D  K  S  S  S  T  A  Y  M  E  L  R  S  L  T  S  E  D  S  S  V
AGGGCCACACTCACCGTGGATAAGTCCAGCAGC
AGGGCCACACTCACCGTGGATAAGTCCAGCAGCACGGCATACATGGAGCTCAGGAGTCTGACCAGCGAAGATAGTTCCGT
TCCCGGTGTGAGTGGCACCTATTCAGGTCGTCGTGCCGTATGTACCTCGAGTCCTCAGACTGGTCGCTTCTATCAAGGCA
TCCCGGTGTG                                                                  GCA

 Y  Y  C  E  S  Q  S  G  A  Y  W  G  Q  G  T  T  V  T  C  S  A  S  G  A  E  F  A
                TAA   TGA   TAA     TCAAGGAACCACAGTCACCGTCTCGGCCTCG
ATACTACTGTGAAAGTCAAAGCGGTGCTTACTGGGGTCAAGGAACCACAGTCACCGTCTCGGCCTCGGGGGCCGAATTCG
TATGATGACACTTTCAGTTTCGCCACGAATGACCCCAGTTCCTTGGTGTCAGTGGCAGAGCCGG
TATGATGACACTTTCAATTTCGACTCGAATTACCCCAGTTCCTTGG

  A  A  A  P  S  V  F  I  F  P  P  S  D  E  Q  L  K  S  G  T  A  S  V  V  C  L
CGGCCGCTGCACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGTTGAAATCTGGAACTGCCTCTGTTGTGTGCCTG
```

Note the change in the nucleotide sequence at H:G106 (in the mo54 stock nucleotide was a T)

5' 3STP_3D5_CDR3for (31bp)

5'- TCAAGGAACCACAGTCACCGTCTCGGCCTCG -3'

GC Content: 61.3 %
Melt Temp: 68.4 ºC
MW: 9426.1 g/mole

Extinction Coeff: 288000 L/(mole·cm)
nmole/OD260: 3.47
µg/OD260: 32.73

3'3STP_3D5_CDR3rev (49bp)
5'- GGT TCC TTG ACC CCA TTA AGC TCA GCT TTA ACT TTC ACA GTA GTA TAC G -3'

GC Content: 42.9 %
Melt Temp: 65.4 ºC
MW: 14956.7 g/mole

Extinction Coeff: 462200 L/(mole·cm)
nmole/OD260: 2.16
µg/OD260: 32.36

5'3STP_3D5_CDR2for (33bp)
5'- AGG GCC ACA CTC ACC GTG GAT AAG TCC AGC AGC -3'

GC Content: 60.6 %
Melt Temp: 69.5 ºC
MW: 10117.6 g/mole

Extinction Coeff: 318900 L/(mole·cm)
nmole/OD260: 3.14
µg/OD260: 31.73

3'3STP_3D5_CDR2rev (51bp)
5'- G TGT GGC CCT GCC CTA GAA CTT TTA GTT TTA GCT CGT GCC GCC ATT ATT CG -3'

GC Content 51.0 %
Melt Temp: 69.6 ºC
MW: 15784.3 g/mole

Extinction Coeff: 509100 L/(mole·cm)
nmole/OD260: 1.96
µg/OD260: 31.00

B.6.2 Library Diversity for 3D5 Heavy Chain 2 and 3

```
G  G  G  S  G  G  G  G  S  S  G  G  G  S  Q  V  Q  L  Q  Q  S  G  P  E  D  V  K
GGTGGTGGTTCTGGCGGCGGCGGCTCCAGTGGTGGTGGATCCCAGGTGCAGCTGCAGCAGTCTGGCCCTGAGGATGTGAA

 P  G  A  S  V  K  I  S  C  K  A  S  G  Y  T  F  T  D  Y  Y  M  N  W  V  K  Q  S
GCCCGGCGCGAGCGTGAAAATCAGTTGTAAAGCCTCTGGATATACATTCACCGATTATTATATGAACTGGGTTAAACAGA

  P  G  K  G  L  E  W  I  G  D  I  N  P  N  N  G  G  T  S  Y  N  Q  K  F  K  G
          GGTCTGGAGTGGATCGGANNSATTHRKHYTNNSRRTNNSNNSNNSNNSTATAACNVSRNKDNBAAGNNS
GTCCAGGAAAGGGTCTGGAGTGGATCGGAGACATTAACCCGAATAATGGCGGCACGAGCTATAACCAAAAGTTCAAGGGC
CAGGTCCTTTCCCAGACCTCACCTAGCCTCTGTAATTGGGCTTATTACCGCCGTGCTCGATATTGGTTTTCAAGTTCCCG
          CCAGACCTCACCTAGCCTNNSTAADYMDRANNSYYANNSNNSNNSNNSATATTGNBSYNMHNVTTCNNS

R  A  T  L  T  V  D  K  S  S  S  T  A  Y  M  E  L  R  S  L  T  S  E  D  S  S  V
AGGGCCACACTCACCGTG                                                          CGT
AGGGCCACACTCACCGTGGATAAGTCCAGCAGCACGGCATACATGGAGCTCAGGAGTCTGACCAGCGAAGATAGTTCCGT
TCCCGGTGTGAGTGGCACCTATTCAGGTCGTCGTGCCGTATGTACCTCGAGTCCTCAGACTGGTCGCTTCTATCAAGGCA
TCCCGGTGTGAGTGGCAC                                                          GCA

 Y  Y  C  E  S  Q           S  G        A        Y  W  G  Q  G  T  T  V  T  C  S
ATACTACTGTGCANVTNNSNNSNNSNNSAGCNNSNNSNNSGCTATGGATTACTGGGGTCAAGG
ATACTACTGTGAAAGTCAA         AGCGGT      GCT      TACTGGGGTCAAGGAACCACAGTCACCGTCT
TATGATGACACTTTCAGTT          TCGCCA      CGA      ATGACCCCAGTTCCTTGGTGTCAGTGGCAGA
TATGATGACACGTNBANNSNNSNNSNNSTCGNNSNNSNNSCGATACCTAATGACCCCAGTTCC

A  S  G  A  E  F  A
CGGCCTCGGGCGCCGAATTCG
```

Note the change in the nucleotide sequence at H:G106 (in the mo54 stock nucleotide was a T)

Heavy CDR2Library (reverse reading frame 5' → 3') - IDEAL
CACGGTGAGTGTGGCCCTSNNCTTVNHMNYSBNGTTATASNNSNNSNNSNNAYYSNNARDMYDAATSNNTCCG ATCCACTCCAGACC

LENGTH: 87
GC CONTENT: 54.4 %

MELT TEMP RANGE:	MIN	MEAN	MAX
	67.1 ºC	72.5 ºC	77.7 ºC

MW: 26668.3 g/mole
Extinc Coeff: 814460 L/(mole·cm)
nmole/OD260: 1.23
µg/OD260: 32.74

Heavy CDR3 Library (reverse reading frame 5' → 3') - IDEAL
CCTTGACCCCAGTAATCCATAGCSNNSNNSNNGCTSNNSNNSNNSNNABNTGCACAGTAGTATACG

LENGTH: 66
GC CONTENT: 54.8 %

MELT TEMP RANGE:	MIN	MEAN	MAX
	66.3 ºC	71.8 ºC	76.4 ºC

MW: 20237.7 g/mole
Extinc Coeff: 626013 L/(mole·cm)
nmole/OD260: 1.60
µg/OD260: 32.33

B.6.3 Rationale for Library Diversity of 3D5 Heavy Chain 2 and 3

Cobaugh, C.W., et al., *Synthetic antibody libraries focused towards peptide ligands.* J Mol Biol, 2008. **378**(3): p. 622-33

Heavy CDR2 Library Diversity					
Kabat #	**aEE**	**3D5**	**CODON**	**Residue Encoded**	**Reason**
50	H	D	NNS	all	Randomization
51	I	I	ATT	I	Common between aEE and 3D5
52	Y	N	HRK	NSRHYW/KQC	Coverage based on Cobaugh (74%) (include a stop codon TAG)
52a		P	HYT	FSLPT/I	Coverage based on Cobaugh (91%) - one primer made without this
53	W	N	NNS	all	Randomization
54	D	N	RRT	GSDN	Coverage based on Cobaugh (89%)
55	D	G	NNS	all	Randomization
56	D	G	NNS	all	Randomization
57	K	T	NNS	all	Randomization
58	R	S	NNS	all	Randomization
59	Y	Y	TAT	Y	Common between aEE and 3D5
60	N	N	AAC	N	Common between aEE and 3D5
61	P	Q	NVS	STCPNQKRHDE/TAYWG	Hydrophillic - Polar Uncharged (only TAG stop codon) by JCPai
62	S	K	RNK	RGTNEVAS/IMKD	Hydrophillic using codon (w/ majority) from Cobaugh
63	L	F	DNB	FYWGALVMIC/STNKRDE	Hydrophobic bias (only TAG stop codon) by JCPai
64	K	K	AAG	K	Common between aEE and 3D5
65	S	G	NNS	all	Randomization

Heavy CDR3 Library Diversity					
Kabat #	**aEE**	**3D5**	**CODON**	**Residue Encoded**	**Reason**
93	A	E	GCA	A	Standard for anti-peptide library (Cobaugh)
94	R	S	NVT	CDGHNPRSTY	charged hydrophillic from sidhu and weiss
95	R	Q	NNS	all	Randomization
96	G		NNS	all	Randomization
97	G		NNS	all	Randomization
98	S		NNS	all	Randomization
99	S	S	AGC	S	Common between aEE and 3D5
100	H	G	NNS	all	Randomization (No Common Characteristics!)
100a	Y		NNS	all	Randomization
100b	Y		TAT	Y	Standard for anti-peptide library (Cobaugh)
100c	A	A	GCT	A	Common between aEE and 3D5
100d	M		ATG	M	Standard for anti-peptide library (Cobaugh)
101	D		GAT(GAC in aEE)	D	Standard for anti-peptide library (Cobaugh)
102	Y	Y	TAC	Y	Common between aEE and 3D5

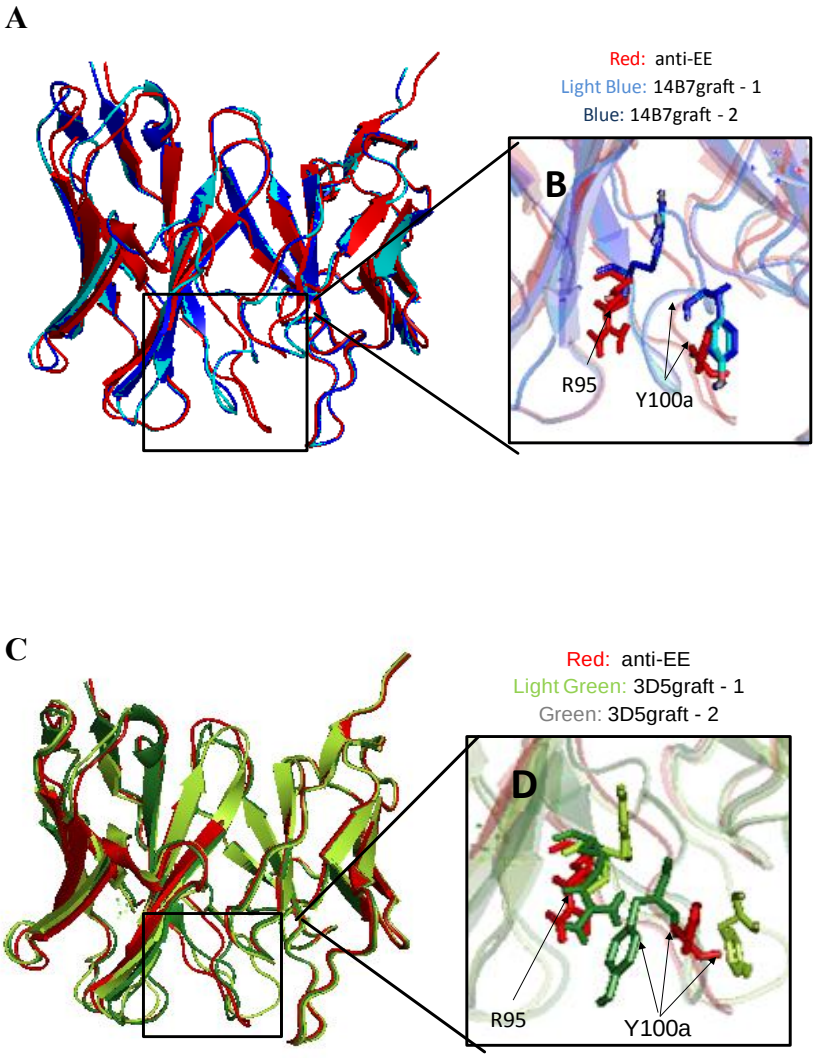
Degenerate Codon Codes:

R –	A		G	
Y –		C		T
M –	A	C		
K –			G	T
S –		C	G	
W –	A			T
H –	A	C		T
B –		C	G	T
V –	A	C	G	
D –	A		G	T
N –	A	C	G	T

B.7 Preliminary Graftwork for anti-peptide scFv variants (3D5g and 14B7g)

B.7.1. Aligned WAM predicted models

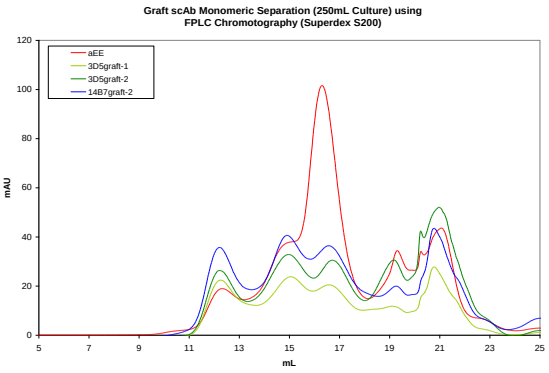
(a) 14B7graft scFv versions 1 and 2 (14B7g-1, 14B7g-2) with αEE scFv, RMSD: 1.60Å (b) enlarged region of CDR3 flexible loops (c) 3D5graft scFv versions 1 and 2 (3D5g-1, 3D5g-2) with αEE scFv, RMSD: 1.04Å.(d)) enlarged region of CDR3 flexible loops



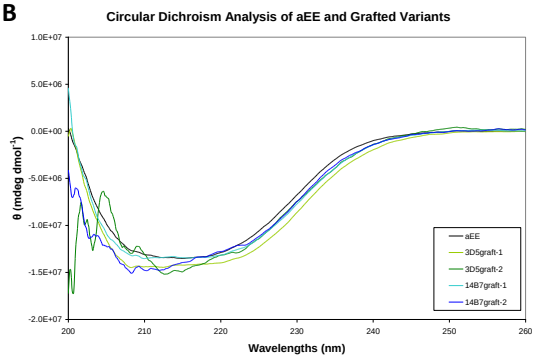
B.7.2 Qualitative Analysis for expression of αEE and engineered grafts

Figure A.8.2. (a) Purification on a size exclusion column (FPLC-SEC) using superdex (S200) column; (b) Circular Dichroism (CD) measurements of αEE and engineered grafts. Table of CD Measurements of grafts with the calculated secondary structure compositions using Chen & Yang reference data set. 14B7graft-1 was not placed on these figures because of it's extremely low-yield.

A



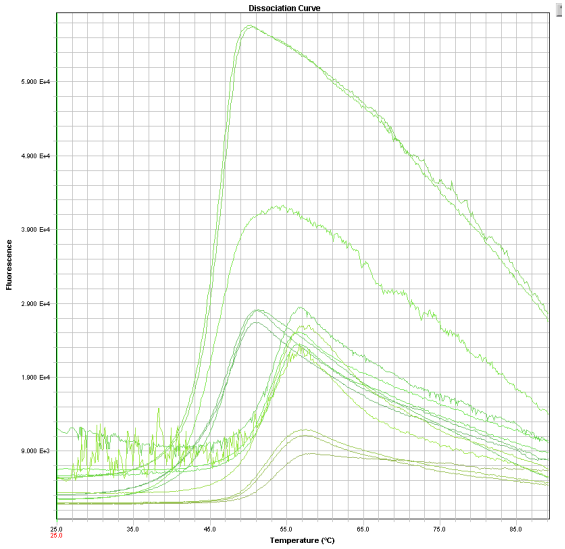
B



	aEE	3D5graft1	3D5graft2	14B7graft1	14B7graft2
% α	21.4	25.3	14	26.8	16.7
% β	37.1	22.2	22.2	35.2	32.3
% Turn	12.3	18.6	22.1	9.6	16.3
% Random	29.3	33.9	41.7	28.4	34.7

B.7.3 Thermal Stability Measurements using from Real-Time PCR

(a) Visualization of the Structural Unfolding (b) Table of Averaged Tm



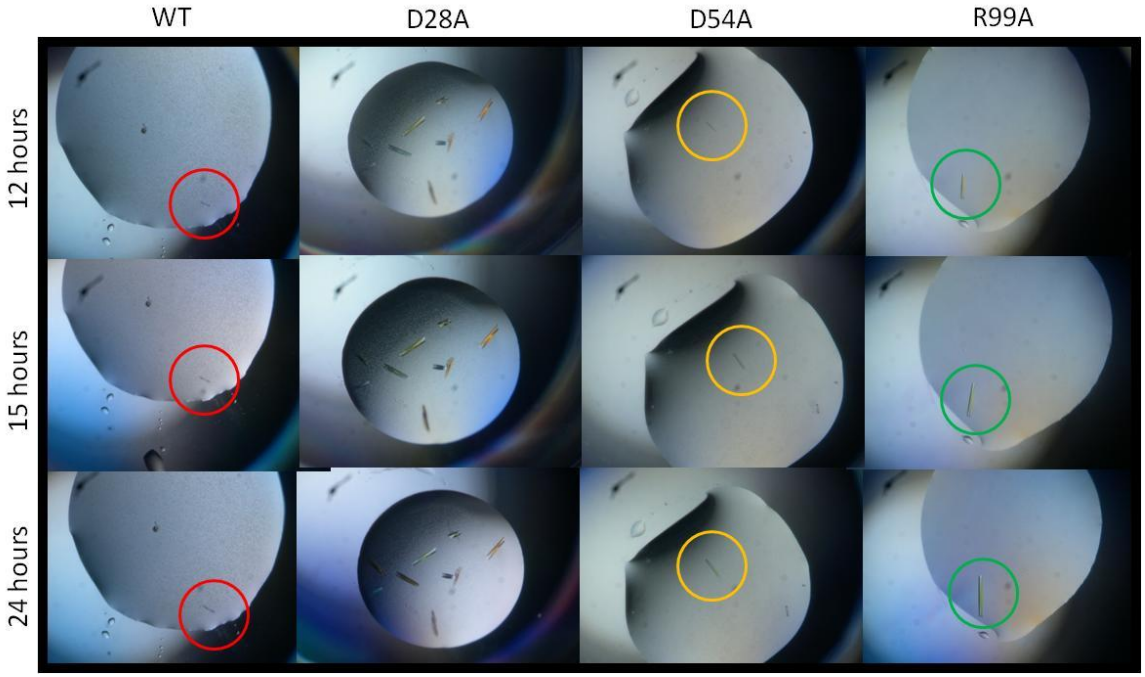
Averaged Tm OVERALL from Real-time PCR		
Protein	Average	Stdev
MBP	62.45	0.152
aEE	53.23	0.379
3D5	46.47	0.451
683	53.64	0.207
3D5graft1A	50.95	0.919
3D5graft1	53.17	0.611
3D5graft2	63.17	0.379
3D5graft48	47.18	0.354
3D5graft48 w H2O	50.95	0.308
3D5graft51	48.03	1.201

B.8 Additional Data for 14B7 Intermolecular Contact Analysis

B.8.1 Table with all the 14B7 Alanines Mutaenesis scFv Variants Engineered

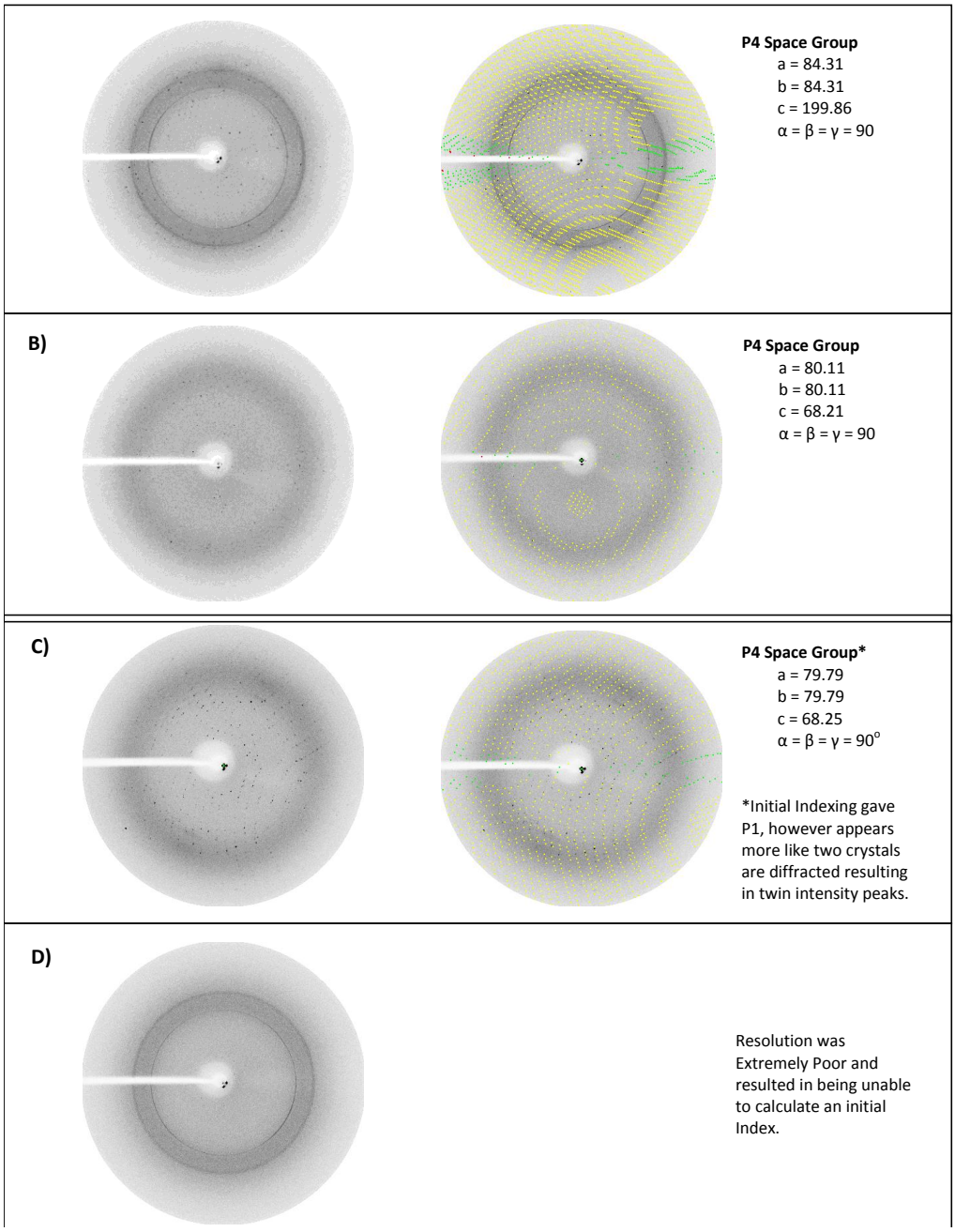
Chain - Kabat #	PDB #	Amino Acid	Chain	ΔΔG (complex) [kcal/mol] in 14B7	ΔΔG (complex) [kcal/mol] in M18	Solvent Accessible Surface Area in 14B7	Surface Area in 14B7 Crystal Contact	% SA of AA in Crystal Contact	% of Crystal Contact (>2% significant)	Solvent Accessible Surface Area in M18	Surface Area in M18 Crystal Contact	Final Crystal Length (um)	CD Analysis (%β sheet)	Total Yield (mg)/250mL Culture	Notes on Crystallization
1	1	D	L		0.08	142.3				68.3/37.5	12.9/67.8				
2	2	I	L		0.32	7.4				7.03	/63.3				
27	27	Q	L		1.45	92.7				100.7/19.7	/80.1				
28	28	D	L	-0.2	0.94	80.8	52.3	0.647277228	0.058770648	84.6/75.2	/0.2	350	69.6	0.79	Very Fast! Cross Shape
30	30	R	L	2.655	0.26	157.5	130.5	0.828571429	0.146645691	119.8/105.4	/50.9	-	87	0.897	X
31	31	N	L	1.13		45	30	0.666666667	0.033711653	41.2/41					
32	32	Y	L	2.17	0.34	73.9	69.7	0.943166441	0.078323407	32.3/68.3	/37.5	500	75.7	0.617	X
49	49	Y	L	0.41		55.4	13.9	0.250902527	0.015619733	25.5/18.9	/12.1				
50	50	Y	L	2.06	0.03	76.1	69	0.906701708	0.077536802	64.3/64.5	0.6/	-	79.3	1.115	X
53	53	R	L	0.22	0.53	122.2	27.7	0.226677578	0.031127093	58.2/135.5	82.4/	400	41.4	0.469	Rod Shape, Normal Rate
67	67	S	L	0		71.8	0.1	0.001392758	0.000112372	73/73.5	39.8/37.9				
92	92	N	L	-0.02	0.07	43.8	11.6	0.264940183	0.013035172	41.7/33.7	/10.4				
93	93	T	L		0.34	61.9				55.1/8.1	/47.4				
30	30	S	H	0	0	48.9				53.2/39.1	/11.2				
31	31	S	H	0.14	-0.17	70	46.7	0.667142857	0.052477806	60.7/8.2	/52.3	417	61.5	0.381	Very Slow! Fanned Rods
33	33	W	H	0.87		17.3	15.5	0.895953757	0.017417687	23.8/9	/19.1	-	37.6	0.518	Not folding properly
52	52	Y	H	0.98	1.23	63.8	58.4	0.915360502	0.065625351	42.1/2	/39.9				
54	55	D	H	0.645	-0.13	112.6	85	0.754884547	0.09551635	10.4/97.6	/10.3	266.67	44.65	0.627	Slighty Faster
56	57	D	H	1.19	-0.34	74	36.3	0.490540541	0.0407911	31.2/64.8	28.6/2.9			0.388	
57	58	T	H		0.36	48.1	1.1	0.022869023	0.001236094	18.3/32.5	20.8/				
64	65	K	H		0.21	152.5	9	0.059016393	0.010113496	72/120.8	69.8/				
97	101	L	H	1.16/0.66	1.54	161.2	131.6	0.816377171	0.147881784	105.3/30.3	/83.3	333.33	74.4	0.655	Rod Shape, Normal Rate
98	102	L	H	1.09	1.29	51	38.9	0.762745098	0.043712777	145.3/43.4	/120.2	-	67.9	0.572	X
99	103	R	H	0.91		133.8	62.6	0.467862481	0.070344983	88.6/68.8	/23.5	416.67	62.1	1.101	Rod Shape, Normal Rate
WT												133.33	73.55	0.934	
T9A+L98A			L,H			193.4	137.3	0.709927611	0.154286999				N/A	0.496	Rod Shape
D28A+D54A	D,D		L,H	0.445		98.1	67.8	0.691131498	0.076188336				N/A		
D28A+W33A	D,W		L,H	0.67									N/A		
D28A+L97A	D,L		L,H	0.96/0.46		131.8	91.2	0.691957511	0.102483425				N/A		Very Fast! Cross Shape
D28A+L98A	D,L		L,H	0.89		214.6	114.9	0.535414725	0.129115631			-	N/A		X
D28A+R99A	D,R		L,H	0.71		270.1	215.5	0.797852647	0.242162041				N/A		Very Fast! Cross Shape
R30A+D54A	R,D		L,H	3.3		208.5	169.4	0.812470024	0.190358467			-	N/A		
R30A+L98A	R,L		L,H	3.745		291.3	193.1	0.662890491	0.216990673			-	N/A		
R30A+R99A	R,R		L,H	3.565		124.9	108.6	0.869495596	0.122036184				N/A		
Y32A+L98A	Y,L		L,H	3.26		207.7	132.3	0.636976408	0.14866839				N/A	1.35	
Y32A+R99A	Y,R		L,H	3.08									N/A	0.054	
Y50A+L97A	Y,L		L,H	3.22/2.72		209.9	131.6	0.626965222	0.147881784				N/A		
Y50A+R99A	Y,L		L,H	2.97		234.8	112.7	0.479982964	0.126643443				N/A	0.413	
R53A+D54A	R,D		L,H	0.865									N/A		
R53A+L97A	R,L		L,H	1.38/0.88		173.2	66.6	0.384526559	0.07482987				N/A		
R53A+L98A	R,L		L,H	1.31		93.4	84.5	0.904710921	0.094954489				N/A		
Y50A+W33A	Y,W		L,H	2.93									N/A		

B.8.2 Crystal Growth Rate of 14B7 Mutants



B.8.3 Representative Diffraction Images from some of the 14B7 variants

Diffracted at the home x-ray source at the University of Texas at Austin. For each variant, the single diffraction image is shown in addition to an image overlaid with the predicted space group symmetry (yellow & green). A) L:D28A B) L:Y32A C) H:R99A D) H: D54A. Note: Due to the low resolution of variant H:D54A, index was unavailable.



B.8.4 Detailed look at 14B7 (D28A) scFv Crystallization Kinetics

Although it was expected that the light chain D28A mutation would improve the crystal contact, it was surprising how drastic the improvement was for crystallization (2 hour crystallization). To better understand how a free energy contribution of < 0 kcal/mol drastically improves crystallization further analysis is required for the D28A mutant. Its distinctive cross shaped crystal formation deviated slightly from the wildtype crystal form, as a result of rapid nucleation on unclean crystal edges. Timed growth rate studies for the D28A variant showed that higher concentrations of scFv produced a biphasic crystallization growth with the transition occurring at 6 hours regardless of the amount of scFv used (Fig. B.1). Crystals diffracted to 2.6Å even with its improved crystallization while retaining a primitive tetragonal index. A full diffraction data set for the D28A crystals is in progress to determine the impact that this substitution has on the interprotein interaction.

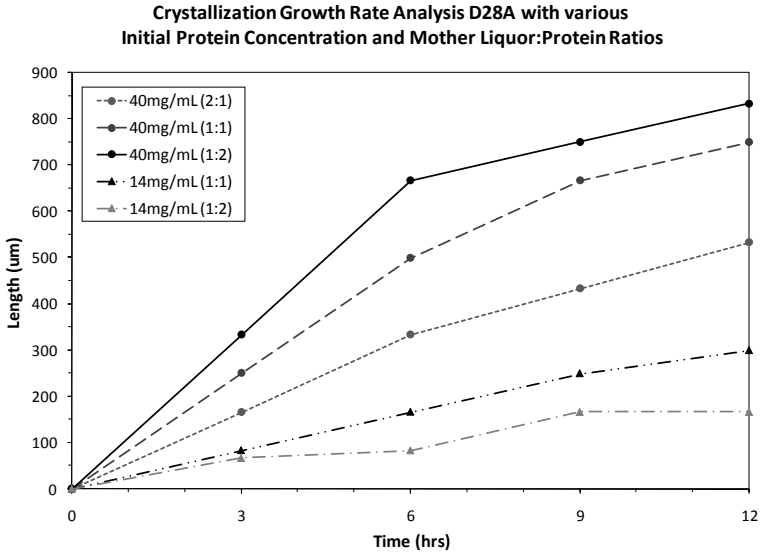


Figure B.1. Timed Crystal Growth Rate Analysis of D28A

Crystallization conditions were performed using various protein concentrations and with different protein:mother liquor ratios.

B.9 Overall protein analysis of crystallization contacts with 3D5 protein family

An anti-peptide antibody fragment family well characterized utilizing crystallization techniques (PDB IDs including 1KTR, 1MAJ, 2CJU, 1DLF, 2UUD, 1DSF, 1WZ1, 1N4X, 2G60) was used as the model group to study common characteristics of protein-protein interactions in crystal lattices. Additionally two scFvs engineered in house derived from the 3D5 scFv were also included in this analysis. Careful analysis of the crystal lattice contacts formed in each of these crystals was examined using *Protein Interfaces, Surfaces and Assemblies* (PISA) (Fig. B.2). The antibodies categorized in this family carry a highly conserved variable light chain with at least ~85% homology to the V_L chain from the murine V_K1 germline with variation occurring in the heavy chain. Although the sequence homology of these fragment antibodies are similar, the diversity in their crystallization conditions and lattices space groups shows their propensity for crystallization (Table B.1, Table B.2). The survey of the antibodies submitted into the *Protein Data Bank* (PDB) in addition to two additional variants engineered in lab displays this family's capability of crystallizing into at least four different space group families (P, C, I, F) and over a wide range of pHs (5.25 – 8.4). A survey of the crystal contacts indicates that the largest ("major") contact involves an approximate area of 444.4 $Å^2$ ± 83.5.

Table B.1 Homology of variants in the 3D5 murine family

Protein	PDB#	Space Group	% Homology to 3D5	% Homology to 3D5 V_L	Major Contact Area ($Å^2$)[a]	Ref.
Anti-histidine 3D5 scFv	1KTR	$P3_221$	100.0	100.0	467.1	[99]
Anti-2-Phenyl-5-Oxazolone NQ16-113.8 scFv	2CJU	$I2_12_12_1$	71.0	95.6	489.2	[102]
Anti-2-Phenyl-5-Oxazolone NNQ10-1.12 scFv	2UUD	C2	69.2	95.6	452.2	[102]
Anti-dansyl monoclonal Fv	1DLF	$P2_12_12_1$	66.1	90.3	407.4	[103]
Anti-dansyl monoclonal Fv	2DLF	$P2_12_12_1$	66.1	90.3	413.8	[103]
Anti-dansyl monoclonal Fv with Ligand	1WZ1	$P4_32_12$	65.2	88.5	325.9	[206]
Anti-cancer Antibody B1 scFv	1DSF	$P6_122$	65.6	85.0	542.6	[207]
scFv – inhibits HIV-1 &HIV-2 proteases	1N4X	$P2_12_12_1$	68.3	91.2	334.4	[208]
Anti-histidine 3D5/His_683	In lab	In progress	92.3	95.6	N/A	Ch. 3
Anti-gluglu peptide tag 3D5/EE_48	3NN8	F23	82.2	95.6	567.3	Ch. 3

[a]Major Contact Area calculated using Protein Interfaces, Surfaces and Assemblies

Table B.2 Crystallization Conditions of select variants in the 3D5 murine family

Protein	Space Group	ML:Protein Ratio (uL)	Protein Conc. (mg/mL)	Crystallization Conditions	Crystal Appearance	Ref.
Anti-histidine 3D5 scFv	$P3_221$	1:5	3.8	0.1M MES (pH 6.4), 0.2M MgAc, 20% PEG8000, 0.02% NaAzide – 4°C	1 weeks	[99]
Anti-2-Phenyl-5-Oxazolone NQ16-113.8 scFv	$I2_12_12_1$	1:1.2	8.0	0.1M Tris-HCL (pH 8.4), 0.2M Ammonium sulfate, 37% PEG 8000, 6% (w/v) sucrose – 17°C	N/A	[102]
Anti-2-Phenyl-5-Oxazolone NNQ10-1.12 scFv	C2	1:1.2	9.0	0.1M Tris-HCl (pH 7.6), 0.2M Ammonium sulfate, 30% (w/v) PEG4000 – 17°C	N/A	[102]
Anti-dansyl monoclonal Fv	$P2_12_12_1$	6:3	5.0	0.025M sodium cacdylate (pH 5.25, 1.6M Ammonium sulfate) – 20°C	3 months	[103]
Anti-dansyl monoclonal Fv	$P2_12_12_1$	6:3	5.0	1.6M Ammonium sulfate and 25mM sodium cacdylate (pH 6.75) – 20°C	3 months	[103]
Anti-dansyl monoclonal Fv with Ligand	$P4_32_12$	1.35:1.35	7.0	0.03M Sodium Cacodylate (pH 6.5), 70mM MgAc, 7% (w/v) PEG 8000	Several weeks	[206]
Anti-cancer Antibody B1 scFv	$P6_122$	3:3	16.0	1.6M Ammonium sulfate, 0.1M HEPES at pH 7.0 – Room Temp, 20°C	1 – 7 days	[207]
scFv – inhibits HIV-1 &HIV-2 proteases	$P2_12_12_1$	1.1	5.0	0.1M sodium citrate (pH 3.5), 17.5% PEG 4000, 0.175M ammonium sulfate	N/A	[208]
Anti-histidine 3D5/His_683	In Progress	1:1 – 1:3	3.8	See table 5.X	4 days	Ch. 3
Anti-gluglu peptide tag 3D5/EE_48	F23	1:1-2	3.8	0.1M Mes (pH 6.4), 0.1 MgAc, 20-24% PEG 8000 – 4°C	4 weeks	Ch. 3

A survey of the residues involved in lattice contact of these similar antibodies showed that a consistent region across the group is involved in one of the crystal lattice contacts even with the unique crystallization condition and space groups. Comparison of amino acids involved in each of the crystal contacts that form the lattice framework for each antibody as been highlighted (Figure 5.1) to determine if there were any consistent driving forces that guide crystallization. Amino acids involved with the largest ("major") crystal contact were colored orange followed by yellow for the second largest contact and green for the third. While some antibodies displayed larger crystal lattice contacts the "conserved contact" was still one of the four major interactions, suggesting that this reoccurring contact is critical to crystallize propensity. The consistent crystal contact

across many of the antibodies (including 3D5, 3D5/EE_48, anti-dansyl Fv and inhibiting scFv for HIV proteases) involves residues in the light chain region (before LCDR1 and between LCDR2 and LCDR3) and contains relatively few amino acids in the heavy CDRs. The residues that are located on the CDRs are shown to be varied among the different members of the family, indicating that these amino acids are not critical for driving the interaction for this crystal contact. Instead these residues appear to contribute as a result of their location association to other critical residues involved in the intermolecular interactions.

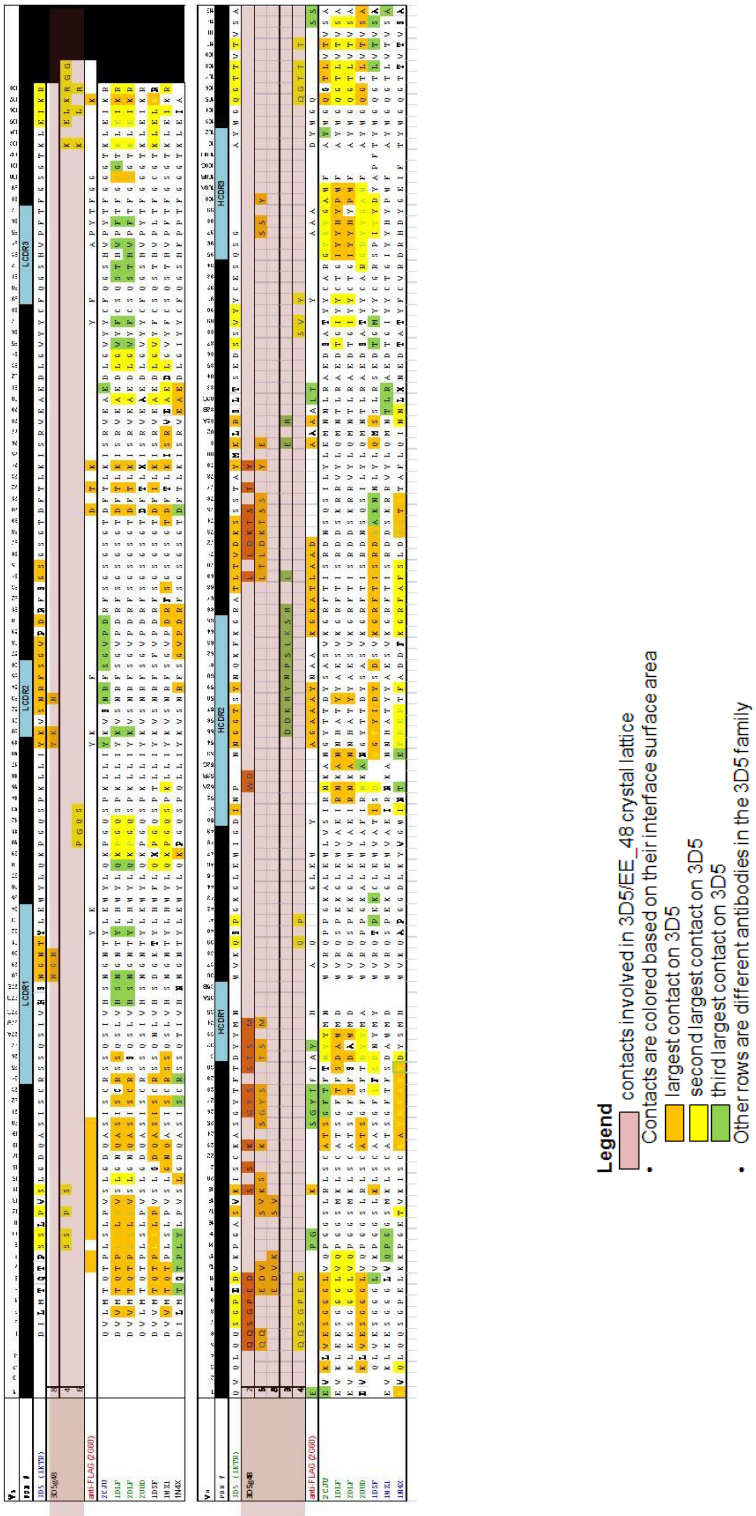


Figure B.2 Contact survey of antibodies with V_L derived from the $V_K1$germline.

Glossary

10XFlag: MDYKAFDNL, improved M2 specific 10x stronger

ccc-DNA: covalently closed circular DNA

CCP: co-crystallization Protein

CD: circular dichroism

CDR: complimentary determining regions

cfu: colony forming units

dU-ssDNA – dUTP-containing single-stranded DNA

EE: EYMPME epitope (peptide)

ELISA: enzyme-linked immunosorbent assay

Fab: fragment antibody

Flag: original tag, DYKDDDK, *see also 10XFlag*

FPLC: fast protein liquid chromatography

HBS: hepes-buffered saline

HCDR: heavy chain CDR (HCDR1, HCDR loop 1, HCDR2, HCDR loop 2; HCDR3, HCDR loop3)

His_6: hexa-histidine

LCDR: light chain CDR (LCDR1, LCDR loop 1, LCDR2, LCDR loop 2; LCDR3, LCDR loop3)

MBP: maltose binding protein

PCR: polymerase chain reaction

PDB: Protein Data Bank

POI: protein of interest

QCM – quick change mutagenesis

REM: random extensive mutagenesis

scFv: single chain antibody fragment (variable)

SEC: size exlusion chromatography

SER: surface entropy reduction

SDM: site directed mutagenesis

SPR: surface plasmon resonance

ssDNA: single stranded DNA

TCR: T-Cell receptor

TMB: tetramethylbenzidine

V_H: variable heavy chain

V_L: variable light chain

References

1. Rhodes, G., *Crystallography Made Clear, Third Edition: A Guide for Users of Macromolecular Models*. 3rd ed. 2006: Academic Press.
2. Elkordy, A.A., R.T. Forbes, and B.W. Barry, *Integrity of crystalline lysozyme exceeds that of a spray-dried form.* Int J Pharm, 2002. **247**(1-2): p. 79-90.
3. Elkordy, A.A., R.T. Forbes, and B.W. Barry, *Stability of crystallised and spray-dried lysozyme.* Int J Pharm, 2004. **278**(2): p. 209-19.
4. Berman, H.M., et al., *The protein structure initiative structural genomics knowledgebase.* Nucleic Acids Res, 2009. **37**(Database issue): p. D365-8.
5. Jimonet, P. and R. Jager, *Strategies for designing GPCR-focused libraries and screening sets.* Curr Opin Drug Discov Devel, 2004. **7**(3): p. 325-33.
6. Schnur, D.M., M.A. Hermsmeier, and A.J. Tebben, *Are target-family-privileged substructures truly privileged?* J Med Chem, 2006. **49**(6): p. 2000-9.
7. Stura, E.A., et al., *Scaffolds for protein crystallisation.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 10 Pt 1): p. 1715-21.
8. Wiencek, J.M., *New strategies for protein crystal growth.* Annu Rev Biomed Eng, 1999. **1**: p. 505-34.
9. Wishart, D., *NMR spectroscopy and protein structure determination: applications to drug discovery and development.* Curr Pharm Biotechnol, 2005. **6**(2): p. 105-20.
10. Chayen, N.E., *Turning protein crystallisation from an art into a science.* Curr Opin Struct Biol, 2004. **14**(5): p. 577-83.
11. Chayen, N.E. and E. Saridakis, *Protein crystallization: from purified protein to diffraction-quality crystal.* Nat Methods, 2008. **5**(2): p. 147-53.
12. Menez, R., et al., *Different crystal packing in Fab-protein L semi-disordered peptide complex.* Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 6): p. 744-9.
13. Caffrey, M., *Membrane protein crystallization.* J Struct Biol, 2003. **142**(1): p. 108-32.
14. Durbin, S.D. and G. Feher, *Protein crystallization.* Annu Rev Phys Chem, 1996. **47**: p. 171-204.
15. Derewenda, Z.S. and P.G. Vekilov, *Entropy and surface engineering in protein crystallization.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 1): p. 116-24.
16. Dasgupta, S., et al., *Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers.* Proteins, 1997. **28**(4): p. 494-514.
17. Janin, J. and F. Rodier, *Protein-protein interaction at crystal contacts.* Proteins, 1995. **23**(4): p. 580-7.
18. Chen, K., L. Kurgan, and M. Rahbari, *Prediction of protein crystallization using collocation of amino acid pairs.* Biochem Biophys Res Commun, 2007. **355**(3): p. 764-9.

19. Smialowski, P., et al., *Will my protein crystallize? A sequence-based predictor.* Proteins, 2006. **62**(2): p. 343-55.
20. McPherson, A., *Protein crystallization in the structural genomics era.* J Struct Funct Genomics, 2004. **5**(1-2): p. 3-12.
21. Wagner, J.R., D.A. Sorgentini, and M.C. Anon, *Relation between solubility and surface hydrophobicity as an indicator of modifications during preparation processes of commercial and laboratory-prepared soy protein isolates.* J Agric Food Chem, 2000. **48**(8): p. 3159-65.
22. Vivar, I., N. Lopez, and A. Musacchio, *Crystallography in the post-genomic era.* Biotecnologia Aplicada, 2007. **24**: p. 209-214.
23. Chothia, C. and J. Janin, *Principles of protein-protein recognition.* Nature, 1975. **256**(5520): p. 705-8.
24. Young, L., R.L. Jernigan, and D.G. Covell, *A role for surface hydrophobicity in protein-protein recognition.* Protein Sci, 1994. **3**(5): p. 717-29.
25. Vekilov, P.G. and A.A. Chernov, *The Physics of Protein Crystallization.* Solid State Physics, 2003. **57**: p. 1-16,16a,17-147.
26. Chernov, A.A., *Protein crystals and their growth.* Journal of Structural Biology, 2003. **142**: p. 3-21.
27. Matsuura, Y. and A.A. Chernov, *Morphology and the strength of intermolecular contacts in protein crystals.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 8): p. 1347-56.
28. Derjaguin, B.V., *Theory of Stability of Colloids and Thin Films*. 1st ed, ed. C. Bureau. 1989, New York: Plenum Publishing Corporation, Springer. 272.
29. Verwey, E.J.W. and J.T.G. Overbeek, *Theory of the stability of lyophobic colloids*. J. Phys. Chem. 1999: Dover Publications. 218.
30. Vekilov, P.G., *What is the Molecular-Level Role of the Solution Components in Protein Crystallization.* Cryst. Growth Des., 2007. **7**(11): p. 2239-2246.
31. Takahashi, D., et al., *Protein-protein interaction on lysozyme crystallization revealed by rotational diffusion analysis.* Biophys J, 2008. **94**(11): p. 4484-92.
32. McQuarrie, D.A., *Statistical mechanics*. 1976, New York: Harper & Row.
33. Deszczynski, M., S.E. Harding, and D.J. Winzor, *Negative second virial coefficients as predictors of protein crystal growth: Evidence from sedimentation equilibrium studies that refutes the designation of those light scattering parameters as osmotic virial coefficients.* Biophysical Chemistry, 2006. **120**: p. 106-113.
34. Jia, Y. and X.Y. Liu, *Prediction of protein crystallization based on interfacial and diffusion kinetics.* Applied Physics Letters, 2005. **87**(10): p. 103902.
35. Asthagiri, D., et al., *A consistent experimental and modeling approach to light-scattering studies of protein-protein interactions in solution.* Biophys J, 2005. **88**(5): p. 3300-9.
36. Haas, C. and J. Drenth, *The interaction energy between two protein molecules related to physical properties of their solution and their crystals and implications for crystal growth.* Journal of Crystal Growth, 1995. **154**(1-2): p. 126-135.

37. Price, W.N., 2nd, et al., *Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data.* Nat Biotechnol, 2009. **27**(1): p. 51-7.
38. Aloy, P. and R.B. Russell, *Interrogating protein interaction networks through structural biology.* Proc Natl Acad Sci U S A, 2002. **99**(9): p. 5896-901.
39. Maynard, J.A., et al., *Protection against anthrax toxin by recombinant antibody fragments correlates with antigen affinity.* Nat Biotechnol, 2002. **20**(6): p. 597-601.
40. Harvey, B.R., et al., *Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from Escherichia coli-expressed libraries.* Proc Natl Acad Sci U S A, 2004. **101**(25): p. 9193-8.
41. Leysath, C.E., et al., *Crystal structure of the engineered neutralizing antibody M18 complexed to domain 4 of the anthrax protective antigen.* J Mol Biol, 2009. **387**(3): p. 680-93.
42. Bergfors, T., *Protein Crystallization: Techniques, Strategies, and Tips*. 1999, La Jolla, California: International University Line.
43. Kendrew, J.C., et al., *Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution.* Nature, 1960. **185**(4711): p. 422-7.
44. Kendrew, J.C., et al., *The species specificity of myoglobin.* Nature, 1954. **174**(4438): p. 946-9.
45. Hansen, C.L., et al., *A microfluidic device for kinetic optimization of protein crystallization and in situ structure determination.* J Am Chem Soc, 2006. **128**(10): p. 3142-3.
46. Anderson, M.J., C.L. Hansen, and S.R. Quake, *Phase knowledge enables rational screens for protein crystallization.* Proc Natl Acad Sci U S A, 2006. **103**(45): p. 16746-51.
47. Vedadi, M., et al., *Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination.* Proc Natl Acad Sci U S A, 2006. **103**(43): p. 15835-40.
48. Neal, B.L., *Why is the osmotic second virial coefficient related to protein crystallization?* Journal of Crystal Growth, 1999. **196**(2-4): p. 377-387.
49. George, A. and W.W. Wilson, *Predicting protein crystallization from a dilute solution property.* Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 4): p. 361-5.
50. Jia, Y. and X.Y. Liu, *From surface self-assembly to crystallization: prediction of protein crystallization conditions.* J Phys Chem B, 2006. **110**(13): p. 6949-55.
51. Lin, Y., et al., *An extensive study of protein phase diagram modification: increasing macromolecular crystallizability by temperature screening.* Cryst. Growth Des., 2008. **8**(12): p. 4277.
52. Saridakis, E. and N.E. Chayen, *Systematic improvement of protein crystals by determining the supersolubility curves of phase diagrams.* Biophys J, 2003. **84**(2 Pt 1): p. 1218-22.
53. Sugiyama, M., et al., *Concentration control for protein crystallization via a continuously-fed crystallization chamber.* Lab Chip, 2008. **8**(8): p. 1398-401.

54. Pechkova, E. and C. Nicolini, *Protein nucleation and crystallization by homologous protein thin film template.* J Cell Biochem, 2002. **85**(2): p. 243-51.
55. Pechkova, E. and C. Nicolini, *Protein nanocrystallography: a new approach to structural proteomics.* Trends Biotechnol, 2004. **22**(3): p. 117-22.
56. Saridakis, E. and N.E. Chayen, *Towards a 'universal' nucleant for protein crystallization.* Trends Biotechnol, 2009. **27**(2): p. 99-106.
57. Dale, G.E., C. Oefner, and A. D'Arcy, *The protein as a variable in protein crystallization.* J Struct Biol, 2003. **142**(1): p. 88-97.
58. D'Arcy, A., et al., *Crystal engineering: a case study using the 24 kDa fragment of the DNA gyrase B subunit from Escherichia coli.* Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 9): p. 1623-5.
59. McElroy, H.E., et al., *Studies on engineering crystallizability by mutation of surface residues of human thymidylate synthase.* Journal of Crystal Growth, 1992. **122**(1-4): p. 265-272.
60. Derewenda, Z.S., *The use of recombinant methods and molecular engineering in protein crystallization.* Methods, 2004. **34**(3): p. 354-63.
61. Rosenbaum, D.M., et al., *GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function.* Science, 2007. **318**(5854): p. 1266-73.
62. Camara-Artigas, A., et al., *Individual interactions influence the crystalline order for membrane proteins.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 9): p. 1281-6.
63. Charron, C., D. Kern, and R. Giege, *Crystal contacts engineering of aspartyl-tRNA synthetase from Thermus thermophilus: effects on crystallizability.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 10 Pt 1): p. 1729-33.
64. Zhang, X.J., J.A. Wozniak, and B.W. Matthews, *Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme.* J Mol Biol, 1995. **250**(4): p. 527-52.
65. Sivozhelezov, V., E. Pechkova, and C. Nicolini, *Mapping electrostatic potential of a protein on its hydrophobic surface: implications for crystallization of Cytochrome P450scc.* J Theor Biol, 2006. **241**(1): p. 73-80.
66. Longenecker, K.L., et al., *Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 5): p. 679-88.
67. Derewenda, Z.S., *Rational protein crystallization by mutational surface engineering.* Structure, 2004. **12**(4): p. 529-35.
68. Cooper, D.R., et al., *Protein crystallization by surface entropy reduction: optimization of the SER strategy.* Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 5): p. 636-45.
69. Nanev, C.N., *How do crystal lattice contacts reveal protein crystallization mechanism.* Crystal Research and Technology, 2008. **43**(9): p. 914-920.
70. Goldschmidt, L., et al., *Toward rational protein crystallization: A Web server for the design of crystallizable protein variants.* Protein Sci, 2007. **16**(8): p. 1569-76.

71. Thornton, J.M. and B.L. Sibanda, *Amino and carboxy-terminal regions in globular proteins.* J Mol Biol, 1983. **167**(2): p. 443-60.
72. Kwong, P.D., et al., *Probability analysis of variational crystallization and its application to gp120, the exterior envelope glycoprotein of type 1 human immunodeficiency virus (HIV-1).* J Biol Chem, 1999. **274**(7): p. 4115-23.
73. Kwong, P.D., et al., *Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody.* Nature, 1998. **393**(6686): p. 648-59.
74. Neau, D.B., et al., *Improving protein crystal quality by selective removal of a Ca(2+)-dependent membrane-insertion loop.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2007. **63**(Pt 11): p. 972-5.
75. Banatao, D.R., et al., *An approach to crystallizing proteins by synthetic symmetrization.* Proc Natl Acad Sci U S A, 2006. **103**(44): p. 16230-5.
76. Yamada, H., et al., *'Crystal lattice engineering,' an approach to engineer protein crystal contacts by creating intermolecular symmetry: crystallization and structure determination of a mutant human RNase 1 with a hydrophobic interface of leucines.* Protein Sci, 2007. **16**(7): p. 1389-97.
77. Ullah, H., et al., *Structure of a signal transduction regulator, RACK1, from Arabidopsis thaliana.* Protein Sci, 2008. **17**(10): p. 1771-80.
78. Keenan, R.J., et al., *DNA shuffling as a tool for protein crystallization.* Proc Natl Acad Sci U S A, 2005. **102**(25): p. 8887-92.
79. Hunte, C. and H. Michel, *Crystallisation of membrane proteins mediated by antibody fragments.* Curr Opin Struct Biol, 2002. **12**(4): p. 503-8.
80. Warke, A. and C. Monmany, *Addressing the protein crystallization bottleneck by cocrystallization.* Cryst. Growth Des., 2007. **7**: p. 2219-2225.
81. Prongay, A.J., et al., *Preparation and crystallization of a human immunodeficiency virus p24-Fab complex.* Proc Natl Acad Sci U S A, 1990. **87**(24): p. 9980-4.
82. Garavito, R.M., D. Picot, and P.J. Loll, *Strategies for crystallizing membrane proteins.* J Bioenerg Biomembr, 1996. **28**(1): p. 13-27.
83. Qutub, Y., et al., *Crystallization of transmembrane proteins in cubo: mechanisms of crystal growth and defect formation.* J Mol Biol, 2004. **343**(5): p. 1243-54.
84. Faham, S., et al., *Crystallization of bacteriorhodopsin from bicelle formulations at room temperature.* Protein Sci, 2005. **14**(3): p. 836-40.
85. Tereshko, V., et al., *Toward chaperone-assisted crystallography: protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (VHH) scaffold.* Protein Sci, 2008. **17**(7): p. 1175-87.
86. Zhou, Y., et al., *Chemistry of ion coordination and hydration revealed by a K+ channel-Fab complex at 2.0 A resolution.* Nature, 2001. **414**(6859): p. 43-8.
87. Fellouse, F.A., C. Wiesmann, and S.S. Sidhu, *Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition.* Proc Natl Acad Sci U S A, 2004. **101**(34): p. 12467-72.

88. Kovari, L.C., C. Momany, and M.G. Rossmann, *The use of antibody fragments for crystallization and structure determinations.* Structure, 1995. **3**(12): p. 1291-3.
89. Ostermeier, C., L.O. Essen, and H. Michel, *Crystals of an antibody Fv fragment against an integral membrane protein diffracting to 1.28 A resolution.* Proteins, 1995. **21**(1): p. 74-7.
90. Binz, H.K. and A. Pluckthun, *Engineered proteins as specific binding reagents.* Curr Opin Biotechnol, 2005. **16**(4): p. 459-69.
91. Huber, T., et al., *In vitro selection and characterization of DARPins and Fab fragments for the co-crystallization of membrane proteins: The Na(+)-citrate symporter CitS as an example.* J Struct Biol, 2007. **159**(2): p. 206-21.
92. Sennhauser, G. and M.G. Grutter, *Chaperone-assisted crystallography with DARPins.* Structure, 2008. **16**(10): p. 1443-53.
93. Cherezov, V., et al., *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor.* Science, 2007. **318**(5854): p. 1258-65.
94. Rasmussen, S.G., et al., *Crystal structure of the human beta2 adrenergic G-protein-coupled receptor.* Nature, 2007. **450**(7168): p. 383-7.
95. Bird, R.E., et al., *Single-chain antigen-binding proteins.* Science, 1988. **242**(4877): p. 423-6.
96. McGregor, D.P., et al., *Spontaneous assembly of bivalent single chain antibody fragments in Escherichia coli.* Mol Immunol, 1994. **31**(3): p. 219-26.
97. Ay, J., et al., *Crystal structure of a phage library-derived single-chain fv fragment complexed with turkey egg-white lysozyme at 2.0 Å resolution.* Journal of Molecular Biology, 2000. **301**(2): p. 239-246.
98. Uysal, S., et al., *Crystal structure of full-length KcsA in its closed conformation.* Proc Natl Acad Sci U S A, 2009. **106**(16): p. 6644-9.
99. Kaufmann, M., et al., *Crystal structure of the anti-His tag antibody 3D5 single-chain fragment complexed to its antigen.* J Mol Biol, 2002. **318**(1): p. 135-47.
100. Lindner, P., et al., *Specific detection of his-tagged proteins with recombinant anti-His tag scFv-phosphatase or scFv-phage fusions.* Biotechniques, 1997. **22**(1): p. 140-9.
101. Muller, K.M., et al., *Tandem immobilized metal-ion affinity chromatography/immunoaffinity purification of His-tagged proteins--evaluation of two anti-His-tag monoclonal antibodies.* Anal Biochem, 1998. **259**(1): p. 54-61.
102. Scotti, C. and E. Gherardi, *Structural basis of affinity maturation of the TEPC15/Vkappa45.1 anti-2-phenyl-5-oxazolone antibodies.* J Mol Biol, 2006. **359**(5): p. 1161-9.
103. Nakasako, M., et al., *The pH-dependent structural variation of complementarity-determining region H3 in the crystal structures of the Fv fragment from an anti-dansyl monoclonal antibody.* J Mol Biol, 1999. **291**(1): p. 117-34.
104. Sidhu, S.S. and G. Weiss, *Constructing phage display libraries by oligonucleotide directed mutagenesis.* 2004.

105. Koide, S. and S.S. Sidhu, *The importance of being tyrosine: lessons in molecular recognition from minimalist synthetic binding proteins.* ACS Chem Biol, 2009. **4**(5): p. 325-34.
106. Grussenmeyer, T., et al., *Complexes of polyoma virus medium T antigen and cellular proteins.* Proc Natl Acad Sci U S A, 1985. **82**(23): p. 7952-4.
107. Ling, M.M. and B.H. Robinson, *Approaches to DNA mutagenesis: an overview.* Anal Biochem, 1997. **254**(2): p. 157-78.
108. Rasila, T.S., M.I. Pajunen, and H. Savilahti, *Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment.* Anal Biochem, 2009. **388**(1): p. 71-80.
109. Arnold, F.H. and G. Georgiou, *Directed Enzyme Evolution*. Methods in Molecular BIology. Vol. 230. 2003, Totowa, New Jersey: Humana Press Inc.
110. Arndt, K.M. and K.M. Muller, *Protein engineering protocols*. Methods in Molecular Biology. Vol. 352. 2007, Totoawa, New Jersey: Humana Press Inc.
111. Bershtein, S. and D.S. Tawfik, *Advances in laboratory evolution of enzymes.* Curr Opin Chem Biol, 2008. **12**(2): p. 151-8.
112. Weiss, G.A., et al., *Rapid mapping of protein functional epitopes by combinatorial alanine scanning.* Proc Natl Acad Sci U S A, 2000. **97**(16): p. 8950-4.
113. Petrounia, I.P. and F.H. Arnold, *Designed evolution of enzymatic properties.* Curr Opin Biotechnol, 2000. **11**(4): p. 325-30.
114. Rubin-Pitel, S.B. and H. Zhao, *Recent advances in biocatalysis by directed enzyme evolution.* Comb Chem High Throughput Screen, 2006. **9**(4): p. 247-57.
115. Hogrefe, H.H., et al., *Creating randomized amino acid libraries with the QuikChange Multi Site-Directed Mutagenesis Kit.* Biotechniques, 2002. **33**(5): p. 1158-60, 1162, 1164-5.
116. Wassman, C.D., et al., *Predicting oligonucleotide-directed mutagenesis failures in protein engineering.* Nucleic Acids Res, 2004. **32**(21): p. 6407-13.
117. Zheng, L., U. Baumann, and J.L. Reymond, *An efficient one-step site-directed and site-saturation mutagenesis protocol.* Nucleic Acids Res, 2004. **32**(14): p. e115.
118. Gratz, A. and J. Jose, *Protein domain library generation by overlap extension (PDLGO): a tool for enzyme engineering.* Anal Biochem, 2008. **378**(2): p. 171-6.
119. Miyazaki, K. and M. Takenouchi, *Creating random mutagenesis libraries using megaprimer PCR of whole plasmid.* Biotechniques, 2002. **33**(5): p. 1033-4, 1036-8.
120. Tseng, W.C., et al., *A novel megaprimed and ligase-free, PCR-based, site-directed mutagenesis method.* Anal Biochem, 2008. **375**(2): p. 376-8.
121. Sanchis, J., et al., *Improved PCR method for the creation of saturation mutagenesis libraries in directed evolution: application to difficult-to-amplify templates.* Appl Microbiol Biotechnol, 2008. **81**(2): p. 387-97.

122. Koyanagi, T., et al., *A rapid, simple, and effective method of constructing a randomly mutagenized plasmid library free from ligation.* Biosci Biotechnol Biochem, 2008. **72**(4): p. 1134-7.
123. Cirino, P.C., K.M. Mayer, and D. Umeno, *Generating mutant libraries using error-prone PCR.* Methods Mol Biol, 2003. **231**: p. 3-9.
124. Clackson, T. and H.B. Lowman, *Phage Display: A Practical Approach (The Practical Approach Series)*. 2004, Oxford, UK: Oxford University Press.
125. Daugherty, P.S., et al., *Development of an optimized expression system for the screening of antibody libraries displayed on the Escherichia coli surface.* Protein Eng, 1999. **12**(7): p. 613-21.
126. Paschke, M., *Phage display systems and their applications.* Appl Microbiol Biotechnol, 2006. **70**(1): p. 2-11.
127. Adda, C.G., et al., *Random sequence libraries displayed on phage: identification of biologically important molecules.* Comb Chem High Throughput Screen, 2002. **5**(1): p. 1-14.
128. Jung, S. and A. Pluckthun, *Improving in vivo folding and stability of a single-chain Fv antibody fragment by loop grafting.* Protein Eng, 1997. **10**(8): p. 959-66.
129. Forrer, P., S. Jung, and A. Pluckthun, *Beyond binding: using phage display to select for structure, folding and enzymatic activity in proteins.* Curr Opin Struct Biol, 1999. **9**(4): p. 514-20.
130. Shivange, A.V., et al., *Advances in generating functional diversity for directed protein evolution.* Curr Opin Chem Biol, 2009. **13**(1): p. 19-25.
131. Carr, P.A. and G.M. Church, *Genome engineering.* Nat Biotechnol, 2009. **27**(12): p. 1151-62.
132. Murase, K., et al., *EF-Tu binding peptides identified, dissected, and affinity optimized by phage display.* Chem Biol, 2003. **10**(2): p. 161-8.
133. Fazelinia, H., P.C. Cirino, and C.D. Maranas, *Extending Iterative Protein Redesign and Optimization (IPRO) in protein library design for ligand specificity.* Biophys J, 2007. **92**(6): p. 2120-30.
134. Shao, Z. and H. Zhao, *DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways.* Nucleic Acids Res, 2009. **37**(2): p. e16.
135. Alper, H., et al., *Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli.* Metab Eng, 2005. **7**(3): p. 155-64.
136. Campbell, E., I.R. Wheeldon, and S. Banta, *Broadening the cofactor specificity of a thermostable alcohol dehydrogenase using rational protein design introduces novel kinetic transient behavior.* Biotechnol Bioeng. **107**(5): p. 763-74.
137. Chin, J.W., et al., *Analysis of NADPH supply during xylitol production by engineered Escherichia coli.* Biotechnol Bioeng, 2009. **102**(1): p. 209-20.
138. Sen, S., V. Venkata Dasu, and B. Mandal, *Developments in directed evolution for improving enzyme functions.* Appl Biochem Biotechnol, 2007. **143**(3): p. 212-23.
139. Wang, C.L., D.C. Yang, and M. Wabl, *Directed molecular evolution by somatic hypermutation.* Protein Eng Des Sel, 2004. **17**(9): p. 659-64.

140. Wong, T.S., et al., *Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution.* Nucleic Acids Res, 2004. **32**(3): p. e26.
141. Hayhurst, A., et al., *Isolation and expression of recombinant antibody fragments to the biological warfare pathogen Brucella melitensis.* J Immunol Methods, 2003. **276**(1-2): p. 185-96.
142. Sutherland, J.N. and J.A. Maynard, *Characterization of a key neutralizing epitope on pertussis toxin recognized by the monoclonal antibody 1B7.* Biochemistry, 2009. **48**(50): p. 11982-11993.
143. Ho, S.N., et al., *Site-directed mutagenesis by overlap extension using the polymerase chain reaction.* Gene, 1989. **77**(1): p. 51-9.
144. Horton, R.M., et al., *Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension.* Gene, 1989. **77**(1): p. 61-8.
145. Kunkel, T.A., *Rapid and efficient site-specific mutagenesis without phenotypic selection.* Proc Natl Acad Sci U S A, 1985. **82**(2): p. 488-92.
146. Stemmer, W.P., *DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution.* Proc Natl Acad Sci U S A, 1994. **91**(22): p. 10747-51.
147. Stemmer, W.P., *Rapid evolution of a protein in vitro by DNA shuffling.* Nature, 1994. **370**(6488): p. 389-91.
148. Rychlik, W., W.J. Spencer, and R.E. Rhoads, *Optimization of the annealing temperature for DNA amplification in vitro.* Nucleic Acids Res, 1990. **18**(21): p. 6409-6412.
149. Pedelacq, J.D., et al., *Engineering soluble proteins for structural genomics.* Nat Biotechnol, 2002. **20**(9): p. 927-32.
150. Yang, J.K., et al., *Directed evolution approach to a structural genomics project: Rv2002 from Mycobacterium tuberculosis.* Proc Natl Acad Sci U S A, 2003. **100**(2): p. 455-60.
151. Wernimont, A. and A. Edwards, *In situ proteolysis to generate crystals for structure determination: an update.* PLoS One, 2009. **4**(4): p. e5094.
152. Hunte, C., et al., *Structure at 2.3 A resolution of the cytochrome bc(1) complex from the yeast Saccharomyces cerevisiae co-crystallized with an antibody Fv fragment.* Structure, 2000. **8**(6): p. 669-84.
153. Milovnik, P., et al., *Selection and characterization of DARPins specific for the neurotensin receptor 1.* Protein Eng Des Sel, 2009. **22**(6): p. 357-66.
154. Byrne, B., et al., *Fusion protein approach to improve the crystal quality of cytochrome bo(3) ubiquinol oxidase from Escherichia coli.* Biochim Biophys Acta, 2000. **1459**(2-3): p. 449-55.
155. Prive, G.G., et al., *Fusion proteins as tools for crystallization: the lactose permease from Escherichia coli.* Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 4): p. 375-9.
156. Fellouse, F.A., et al., *High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries.* J Mol Biol, 2007. **373**(4): p. 924-40.

157. Prickett, K.S., D.C. Amberg, and T.P. Hopp, *A calcium-dependent antibody for identification and purification of recombinant proteins.* Biotechniques, 1989. **7**(6): p. 580-9.
158. Krebber, A., et al., *Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system.* J Immunol Methods, 1997. **201**(1): p. 35-55.
159. Cobaugh, C.W., et al., *Synthetic antibody libraries focused towards peptide ligands.* J Mol Biol, 2008. **378**(3): p. 622-33.
160. Maynard, J., et al., *High-level bacterial secretion of single-chain alphabeta T-cell receptors.* J Immunol Methods, 2005. **306**(1-2): p. 51-67.
161. Sambrook, J., *Molecular Cloning*. 3rd ed. 2001: Cold Spring Harbor Laboratory Press.
162. Lavinder, J.J., et al., *High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering.* J Am Chem Soc, 2009. **131**(11): p. 3794-5.
163. Kabsch, W., *Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants.* J Applied Cryst, 1993. **26**: p. 795-800.
164. *The CCP4 suite: programs for protein crystallography.* Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 5): p. 760-3.
165. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
166. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10037-41.
167. Sanner, M.F., *Python: a programming language for software integration and development.* J Mol Graph Model, 1999. **17**(1): p. 57-61.
168. Comeau, S.R., et al., *ClusPro: a fully automated algorithm for protein-protein docking.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W96-9.
169. Baca, M., et al., *Antibody humanization using monovalent phage display.* J Biol Chem, 1997. **272**(16): p. 10678-84.
170. Reiersen, H. and A.R. Rees, *The hunchback and its neighbours: proline as an environmental modulator.* Trends Biochem Sci, 2001. **26**(11): p. 679-84.
171. Xu, J.L. and M.M. Davis, *Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities.* Immunity, 2000. **13**(1): p. 37-45.
172. de Haard, H.J., et al., *Absolute conservation of residue 6 of immunoglobulin heavy chain variable regions of class IIA is required for correct folding.* Protein Eng, 1998. **11**(12): p. 1267-76.
173. Chothia, C. and A.M. Lesk, *Canonical structures for the hypervariable regions of immunoglobulins.* J Mol Biol, 1987. **196**(4): p. 901-17.
174. Rothlisberger, D., K.M. Pos, and A. Pluckthun, *An antibody library for stabilizing and crystallizing membrane proteins - selecting binders to the citrate carrier CitS.* FEBS Lett, 2004. **564**(3): p. 340-8.

175. Veesler, D., et al., *Crystal structure and function of a DARPin neutralizing inhibitor of lactococcal phage TP901-1: comparison of DARPin and camelid VHH binding mode.* J Biol Chem, 2009. **284**(44): p. 30718-26.
176. Rader, C., D.A. Cheresh, and C.F. Barbas, 3rd, *A phage display approach for rapid antibody humanization: designed combinatorial V gene libraries.* Proc Natl Acad Sci U S A, 1998. **95**(15): p. 8910-5.
177. Derewenda, Z.S., *Application of protein engineering to enhance crystallizability and improve crystal properties.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 5): p. 604-15.
178. Cuello, L.G., et al., *Structural mechanism of C-type inactivation in K(+) channels.* Nature. **466**(7303): p. 203-8.
179. Koide, S., *Engineering of recombinant crystallization chaperones.* Curr Opin Struct Biol, 2009. **19**(4): p. 449-57.
180. Slootstra, J.W., et al., *Identification of new tag sequences with differential and selective recognition properties for the anti-FLAG monoclonal antibodies M1, M2 and M5.* Mol Divers, 1997. **2**(3): p. 156-64.
181. Dueber, J.E., et al., *Synthetic protein scaffolds provide modular control over metabolic flux.* Nat Biotechnol, 2009. **27**(8): p. 753-9.
182. Granier, S., et al., *Structure and conformational changes in the C-terminal domain of the beta2-adrenoceptor: insights from fluorescence resonance energy transfer studies.* J Biol Chem, 2007. **282**(18): p. 13895-905.
183. Malhotra, A., *Tagging for protein expression.* Methods Enzymol, 2009. **463**: p. 239-58.
184. Ki, J.J., et al., *A periplasmic fluorescent reporter protein and its application in high-throughput membrane protein topology analysis.* J Mol Biol, 2004. **341**(4): p. 901-9.
185. Cieslik, M. and Z.S. Derewenda, *The role of entropy and polarity in intermolecular contacts in protein crystals.* Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 5): p. 500-9.
186. Pellicane, G., G. Smith, and L. Sarkisov, *Molecular dynamics characterization of protein crystal contacts in aqueous solutions.* Phys Rev Lett, 2008. **101**(24): p. 248102.
187. Janin, J., *Specific versus non-specific contacts in protein crystals.* Nat Struct Biol, 1997. **4**(12): p. 973-4.
188. Carugo, O. and P. Argos, *Protein-protein crystal-packing contacts.* Protein Sci, 1997. **6**(10): p. 2261-3.
189. Brewer, S.J., et al., *Engineering proteins to enable their isolation in a biologically active form.* Bioprocess Technol, 1991. **12**: p. 239-66.
190. Sassenfeld, H.M., *Engineering proteins for purification.* Trends Biotechnol, 1990. **8**(4): p. 88-93.
191. Jancarik, J., et al., *Crystallization and preliminary X-ray diffraction study of the ligand-binding domain of the bacterial chemotaxis-mediating aspartate receptor of Salmonella typhimurium.* J Mol Biol, 1991. **221**(1): p. 31-4.

192. Trevino, S.R., J.M. Scholtz, and C.N. Pace, *Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa.* J Mol Biol, 2007. **366**(2): p. 449-60.
193. Robien, M.A., et al., *An improved crystal form of Plasmodium falciparum peptide deformylase.* Protein Sci, 2004. **13**(4): p. 1155-63.
194. Kortemme, T., D.E. Kim, and D. Baker, *Computational alanine scanning of protein-protein interfaces.* Sci STKE, 2004. **2004**(219): p. pl2.
195. Kortemme, T. and D. Baker, *A simple physical model for binding energy hot spots in protein-protein complexes.* Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14116-21.
196. Lee, B. and F.M. Richards, *The interpretation of protein structures: estimation of static accessibility.* J Mol Biol, 1971. **55**(3): p. 379-400.
197. Chothia, C., et al., *Conformations of immunoglobulin hypervariable regions.* Nature, 1989. **342**(6252): p. 877-83.
198. Jirholt, P., et al., *A central core structure in an antibody variable domain determines antigen specificity.* Protein Eng, 2001. **14**(1): p. 67-74.
199. Chen, Y.H. and J.T. Yang, *A new approach to the calculation of secondary structures of globular proteins by optical rotatory dispersion and circular dichroism.* Biochem Biophys Res Commun, 1971. **44**(6): p. 1285-91.
200. Leslie, A.G.W., *Recent changes to the MOSFLM package for processing film and image plate data.* CCP4 and ESF-EACMB Newsletter on Protein Crystallography, 1991.
201. CCP4 and C.C.P.N. 4, *The CCP4: suite: programs for protein crystallography.* Acta Cryst., 1994. **D50**: p. 760-763.
202. Dale, G.E., et al., *Improving protein solubility through rationally designed amino acid replacements: solubilization of the trimethoprim-resistant type S1 dihydrofolate reductase.* Protein Eng, 1994. **7**(7): p. 933-9.
203. Zhang, F., et al., *Crystal structure of the obese protein leptin-E100.* Nature, 1997. **387**(6629): p. 206-9.
204. Patel, S.B., et al., *Lattice stabilization and enhanced diffraction in human p38 alpha crystals by protein engineering.* Biochim Biophys Acta, 2004. **1696**(1): p. 67-73.
205. Derewenda, U., et al., *The structure of Yersinia pestis V-antigen, an essential virulence factor and mediator of immunity against plague.* Structure, 2004. **12**(2): p. 301-6.
206. Nakasako, M., et al., *Conformational dynamics of complementarity-determining region H3 of an anti-dansyl Fv fragment in the presence of its hapten.* J Mol Biol, 2005. **351**(3): p. 627-40.
207. Almog, O., et al., *Crystal structure of the disulfide-stabilized Fv fragment of anticancer antibody B1: conformational influence of an engineered disulfide bond.* Proteins, 1998. **31**(2): p. 128-38.

208. Lescar, J., et al., *Structure of a single-chain Fv fragment of an antibody that inhibits the HIV-1 and HIV-2 proteases.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 5): p. 955-7.

Vita

Jennifer Chentzu Pai attended Ward Melville High School in Setauket, New York where she graduated in the 3rd percentile in 2002. Her interest in the engineering sciences led to her enrollment at Cornell University in Ithaca NY, where she received her Bachelor of Science in Chemical Engineering in 2006. As she pursued her undergraduate studies, she was a Summer Undergraduate Laboratory Intern (SULI) under Dr. Stephen Dewey at the Brookhaven National Lab in Upton, NY during 2005 and 2006. Upon graduation, she was immediately accepted into the Chemical Engineering Department at the University of Texas at Austin to pursue a doctoral degree in Chemical Engineering. While at the University of Texas, she obtained her Masters in Science in Chemical Engineering in the Spring of 2010.

Email address: jennifer.c.pai@gmail.com

This dissertation was typed by Jennifer C. Pai